

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-187725

(43)Date of publication of application : 21.07.1998

(51)Int.Cl.

G06F 17/27

G06F 17/24

(21)Application number : 08-348586

(71)Applicant : NEC CORP

(22)Date of filing : 26.12.1996

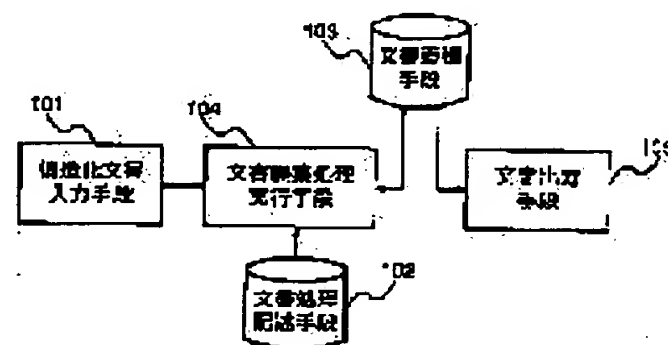
(72)Inventor : OOTSUBO MOTOHIDE

## (54) DOCUMENT EDITING SYSTEM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To extract a partial structure from a structured document while maintaining connection between extracted places according to a pattern description that can specify extraction places as units if the structure of a document when necessary partial structures are extracted from structured documents and put together to generate a new structured document.

**SOLUTION:** This system has a structured document input means 101, which reads in a structured document, a document process describing means 102 where editing place descriptions for the structure document are given, a document storage means 103, which divides the partial structures of the structured document by quotation specifies and store then in the order of matches with the structured document, a document editing process execution means 104, which carries out matching process on the editing place descriptions and the structure document, and extracts and stores the partial structures specified with the quotation specifies in the document storage means, and a document output means 105, which outputs the document stored in the document storage means as it is or after processing it.



## LEGAL STATUS

[Date of request for examination]	26.12.1996
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3156613
[Date of registration]	09.02.2001
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

Copyright (C); 1998,2003 Japan Patent Office



【書類名】 特許願

【整理番号】 33508385SH

【提出日】 平成8年12月26日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/24

【発明の名称】 文書編集方式

【請求項の数】 19

【発明者】

【住所又は居所】 東京都港区芝五丁目7番1号  
本電気株式会社内

日

【氏名】 大坪 基秀

【特許出願人】

【識別番号】 000004237

【氏名又は名称】 日本電気株式会社

【代表者】 金子 尚志

【代理人】

【識別番号】 100082935

【弁理士】

【氏名又は名称】 京本 直樹

【電話番号】 03-3454-1111

【選任した代理人】

【識別番号】 100082924

【弁理士】

【氏名又は名称】 福田 修一

【電話番号】 03-3454-1111

【選任した代理人】

【識別番号】 100085268

【弁理士】

【氏名又は名称】 河合 信明

【電話番号】 0 3 - 3 4 5 4 - 1 1 1 1

【手数料の表示】

【予納台帳番号】 0 0 8 2 7 9

【納付金額】 2 1 0 0 0

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9 1 1 5 6 9 9

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書編集方式

【特許請求の範囲】

【請求項 1】 複数の構造化文書を読み込み、  
文書の構造を単位として複数の抽出箇所を指定するパターン記述に従って、抽出箇所間のつながりを維持しながら、前記構造化文書から部分構造を抽出することを特徴とする文書編集方式。

【請求項 2】 1 つ以上の構造化文書を順次読み込む構造化文書入力手段と、  
前記構造化文書入力手段の読み込んだ入力文書の一部である部分構造を指定する、文書タグ、文書文字パターン、否定指定子、引用指定子、順序接続子、階層接続子、括弧によって構成される編集箇所記述要素を、AND 接続子、OR 接続子、括弧で結合した編集箇所記述が 1 つ以上記述されている文書処理記述手段と、  
前記部分構造を、前記文書処理記述手段に記述された編集箇所記述中の引用指定子毎に分け、かつ、前記入力文書と前記編集箇所記述とのマッチが成立した順に蓄積する文書蓄積手段と、  
前記文書処理記述手段から編集箇所記述を取り出し、前記編集箇所記述を構成する各編集箇所記述要素と入力文書とのマッチを試み、マッチが成立した前記編集箇所記述中に記述された引用指定子に対応する部分構造を前記文書蓄積手段に蓄積し、また、前記編集箇所記述中の AND 接続子で結合された編集箇所記述要素の少なくとも 1 つのマッチが成立しない場合には、前記 AND 接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除し、また、前記編集箇所記述中の OR 接続子で結合された編集箇所記述要素のどのマッチも成立しない場合には、前記 OR 接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除して抽出処理を行う文書編集処理実行手段と、  
前記文書蓄積手段に蓄積された文書をそのまま、あるいは加工して出力する文書出力手段と

を有することを特徴とする文書編集方式。

【請求項 3】 1 つ以上の構造化文書を順次読み込む構造化文書入力手段と、  
前記構造化文書入力手段の読み込んだ入力文書の一部である部分構造を指定する、文書タグ、文書文字パターン、ワイルドカードタグ、否定指定子、引用指定子、順序接続子、階層接続子、括弧によって構成される編集箇所記述要素を、AND 接続子、OR 接続子、括弧で結合した編集箇所記述が 1 つ以上記述されている文書処理記述手段と、  
前記部分構造を、前記文書処理記述手段に記述された編集箇所記述中の引用指定子毎に分け、かつ、前記入力文書と前記編集箇所記述とのマッチが成立した順に蓄積する文書蓄積手段と、  
前記文書処理記述手段から編集箇所記述を取り出し、前記編集箇所記述を構成する各編集箇所記述要素と入力文書とのマッチを試み、マッチが成立した前記編集箇所記述中に記述された引用指定子に対応する部分構造を前記文書蓄積手段に蓄積し、また、前記編集箇所記述中の AND 接続子で結合された編集箇所記述要素の少なくとも 1 つのマッチが成立しない場合には、前記 AND 接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除し、また、前記編集箇所記述中の OR 接続子で結合された編集箇所記述要素のどのマッチも成立しない場合には、前記 OR 接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除して抽出処理を行う文書編集処理実行手段と、  
前記文書蓄積手段に蓄積された文書をそのまま、あるいは加工して出力する文書出力手段と  
を有することを特徴とする文書編集方式。

【請求項 4】 前記文書処理記述手段は、前記文書編集処理実行手段の前記抽出処理の際に実行する動作を記述した動作記述を前記編集箇所記述に対応付けて記述した編集処理記述が 1 つ以上記述され、  
前記文書編集処理実行手段は、前記抽出処理を実行して、前記文書蓄積手段中に残った抽出した部分構造と前記動作記述とを前記動作記述解釈手段へ通知し、

前記文書出力手段は、出力文書蓄積手段に蓄積された構造化文書や部分構造をそのまま、あるいは加工して出力し、

構造化文書や部分構造を蓄積する出力文書蓄積手段と、

前記動作記述を解釈し、文書の追加、文書の結合、文書数のカウントなどの編集処理を実行して、構造化文書や部分構造を前記出力文書蓄積手段に蓄積する動作記述解釈手段と

をさらに有することを特徴とする請求項 2 または 3 記載の文書編集方式。

【請求項 5】 前記読み込まれた各文書に対しての前記抽出処理を開始する前に毎回実行する動作を記述した文書毎前処理記述手段と、

前記読み込まれた各文書に対しての前記編集処理が完了する度に毎回実行する動作を記述した文書毎後処理記述手段と、

前記読み込まれた各文書に対しての前記抽出処理が行われる前に、毎回、前記文書毎前処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する文書毎前処理実行手段と、

前記読み込まれた各文書に対しての前記編集処理が完了する度に、毎回、前記文書毎後処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する文書毎後処理実行手段と

をさらに有することを特徴とする請求項 4 記載の文書編集方式。

【請求項 6】 前記構造化文書入力手段が最初の文書を読み込む前に実行する動作を記述した初期処理記述手段と、

前記読み込まれた全ての文書に対する前記抽出処理および前記編集処理が完了した後に実行する動作を記述した終了時処理記述手段と、

前記構造化文書入力手段が最初の文書を読み込む前に、前記初期処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する初期処理実行手段と、

前記読み込まれた全ての文書に対する前記抽出処理および前記編集処理が完了した後に、前記終了時処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する終了時処理実行手段と

をさらに有することを特徴とする請求項 4 または 5 記載の文書編集方式。

【請求項 7】 構造化文書の構造名（以降、タグ名と呼ぶ）変更指示を受け

た場合には、指定された文書中の指定されたタグ名を指定された別のタグ名に置き換え、また、構造削除指示を受けた場合には、指定された文書中の指定されたタグ名から始まる構造を削除し、また、文字列変更指示を受けた場合には、指定された文書中の指定された文字列を指定された別の文字列に置き換え、また、文字列削除指示を受けた場合には、指定された文書中の指定された文字列を削除する文書修正手段をさらに有し、

前記動作記述解釈手段は、動作記述の解釈に応じて文書修正手段への指示を含む編集処理を実行することを特徴とする

請求項 4、5 または 6 記載の文書編集方式。

【請求項 8】 指定された構造化文書や部分構造を指定されたキーに従ってソートする文書ソート手段をさらに有し、

前記動作記述解釈手段は、動作記述の解釈に応じて文書ソート手段への指示を含む編集処理を実行することを特徴とする

請求項 4、5、6 または 7 記載の文書編集方式。

【請求項 9】 前記構造化文書入力手段は、構造化文書を読み込む際に文書構造をチェックし、

前記文書出力手段は、文書を出力する際に文書構造をチェックすることを特徴とする

請求項 4、5、6、7 または 8 記載の文書編集方式。

【請求項 10】 前記文書蓄積手段は、抽出した構造化文書毎に部分構造を分け、さらに前記文書処理記述手段に記述された編集箇所記述中の引用指定子毎に分けて蓄積することを特徴とする

請求項 4、5、6、7、8 または 9 記載の文書編集方式。

【請求項 11】 前記読み込まれた文書の中の 1 つもしくは複数の文書に対する削除要求を受け付け、前記削除要求を前記文書編集処理実行手段および前記動作記述解釈手段に通知する入力文書削除要求受信手段をさらに有し、

前記文書編集処理実行手段は、前記入力文書削除要求受信手段から削除要求があった場合には、前記削除要求によって指定された入力文書に関する部分構造を前記文書蓄積手段から削除し、

前記動作記述解釈手段は、前記入力文書削除要求受信手段から削除要求があった場合には、前記削除要求によって指定された入力文書に関する部分構造を前記出力文書蓄積手段から削除することを特徴とする

請求項 10 記載の文書編集方式。

【請求項 12】 前記文書編集処理実行手段は、前記読み込まれた各文書と前記各編集箇所記述要素とのマッチ処理を行う際に、前記文書の任意の部分構造を始点として前記編集箇所記述要素とマッチする全ての部分構造を求めて前記抽出処理を実行することを特徴とする

請求項 2、3、4、5、6、7、8、9、10 または 11 記載の文書編集方式。

【請求項 13】 前記文書編集処理実行手段は、前記読み込まれた各文書と前記各編集箇所記述要素とのマッチ処理を行う際に、前記文書の任意の部分構造を始点として前記編集箇所記述要素とマッチする最初の部分構造を求めて前記抽出処理を実行することを特徴とする

請求項 2、3、4、5、6、7、8、9、10 または 11 記載の文書編集方式。

【請求項 14】 前記文書編集処理実行手段は、前記読み込まれた各文書と前記各編集箇所記述要素とのマッチ処理を行う際に、前記文書の任意の部分構造を始点として前記編集箇所記述要素とマッチする最後の部分構造を求めて前記抽出処理を実行することを特徴とする

請求項 2、3、4、5、6、7、8、9、10 または 11 記載の文書編集方式。

【請求項 15】 前記文書編集処理実行手段は、前記読み込まれた各文書と前記各編集箇所記述要素とのマッチ処理を行う際に、前記文書の任意の部分構造を始点として前記編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も外側の部分構造を選択して前記抽出処理を実行することを特徴とする

請求項 2、3、4、5、6、7、8、9、10 または 11 記載の文書編集方式。

【請求項 16】 前記文書編集処理実行手段は、前記読み込まれた各文書と前記各編集箇所記述要素とのマッチ処理を行う際に、前記文書の任意の部分構造を始点として前記編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も内側の部分構造を選択して前記抽出処理を実行することを特徴

とする

請求項 2、3、4、5、6、7、8、9、10 または 11 記載の文書編集方式。

【請求項 17】 前記構造化文書入力手段は、入力文書が与えられる度に読み込むことを特徴とする

請求項 2、3、4、5、6、7、8、9、10、11、12、13、14、15  
または 16 記載の文書編集方式。

【請求項 18】 前記文書出力手段は、全ての入力文書に対する文書編集処理が完了した後に出力処理を行うことを特徴とする

請求項 2、3、4、5、6、7、8、9、10、11、12、13、14、15  
、16 または 17 記載の文書編集方式。

【請求項 19】 前記文書出力手段は、1つの入力文書に対する文書編集処理が完了する度に出力処理を行うことを特徴とする

請求項 2、3、4、5、6、7、8、9、10、11、12、13、14、15  
、16 または 17 記載の文書編集方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は文書編集方式に関し、特に、例えば、個人単位で作成した作業進捗報告などの報告書をまとめてグループ全体の報告書や一覧表を作成する際に、各人の「目標」や「進捗」や「成果」などの部分構造を抽出して、それぞれまとめる作業を自動的に行うための文書編集方式に関するものである。

【0002】

また、例えば、ある1つの製品の各部分を別々の拠点もしくは別々のグループが設計、作成する分散開発において、各拠点もしくは各グループが設計、作成した部分に関する仕様書などのドキュメントをまとめて、製品の仕様書を作成する際の作業を自動化するための文書編集方式に関するものである。

【0003】

さらに、例えば、前記作業進捗報告の例や前記製品仕様書作成の例において、各人、各拠点もしくは各グループが作成した文書が、電子メールなどによってあ

る一箇所に送付される場合に、電子メールが到着する度に、それまで到着した文書とともに構造毎にまとめあげて出力文書を自動的に更新する処理や、以前に送付したメールに対する削除要求を処理して出力文書を自動的に更新するための文書編集方式に関するものである。

#### 【 0 0 0 4 】

なお、前記報告書や仕様書など、本発明の文書編集方式の入力となる文書は、あらかじめ決められた記号等で文書構造を明確に表した S G M L (Standard Generalized Markup Language) に代表される構造化文書であることを前提としており、特に以降の具体的説明では、S G M L 形式の文書を扱う例を用いる。S G M L に関しては、例えば、オーム社刊、吉岡誠 編著の” S G M L のススメ”などの書籍に詳述されている。

#### 【 0 0 0 5 】

##### 【従来の技術】

従来の文書編集方式は、複数の文書の必要箇所を抽出し加工し 1 つの文書にまとめるための技術には、次の 2 種類があった。

(1) 例えば U N I X オペレーティングシステムに搭載された s e d コマンドや a w k コマンドなどのように、文書を構造のない文字列の集合と見なして、文字パターンとのマッチングによって文書の必要箇所を抽出する文書編集方式があった。

(2) 特開平 6 - 2 5 9 4 2 1 号公報などに代表される、” 節”、” 表題” などの単純文字列パターン（本明細書の請求項 2 での「文書タグ」に相当する）と、包含関係を表す” / ”（本明細書の請求項 2 での「階層接続子」に相当する）や順序関係を表す” # ”（本明細書の請求項 2 での「順序接続子」に相当する）などの接続表現記号、” ( ” や ” ) ” などの括弧（本明細書の請求項 2 での「括弧」に相当する）、等の構文要素からなる文書構造パターンを用いて、文書構造を保って必要箇所を抽出・加工する文書編集方式があった。

#### 【 0 0 0 6 】

##### 【発明が解決しようとする課題】

上述した従来の文書編集方式は、

(1) 前記文書を構造のない文字列の集合と見なして処理する従来の文書編集方式には、例えば、特開平6-259421号公報の「従来の技術」の(6)および「発明が解決しようとする課題」などでも指摘している通り、文書の構造の意味を解釈せずに文字列単位で照合しているため、文書の部分構造のマッチングや抽出といった文書構造単位での処理が困難であるという問題点があり、文書の構造を単位としてパターン記述し、文書編集するための手段が必要となった。

#### 【0007】

(2) また、特開平6-259421号公報に代表される従来の文書編集方式には、以降に述べる(2-1)および(2-2)および(2-3)のような問題点があった。

#### 【0008】

(2-1) 特開平6-259421号公報に代表される従来の文書編集方式のパターン記述では、パターン記述の構文要素に、文書の抽出箇所を示す引用指定子がないため、1つのパターンによって抽出することのできる部分構造は1種類に限られる。すなわち、1つの文書から複数種類の部分構造を抽出したい場合には、抽出したい部分構造の種類だけパターン記述を用意して、別々に部分構造を抽出することになる。しかしながら、各パターン記述の間の関係までは規定できないため、例えば、各パターン記述から抽出した部分構造から、部分構造間の関係を問わずに部分構造の組を作成して出力文書を作成するような文書編集には適用できるが、ある位置関係を満たす部分構造の組を抽出して出力文書を作成するような文書編集には適用できない、といった問題があった。

#### 【0009】

例を用いて前記(2-1)の問題点を説明する。例えば、各個人や小グループから文書として送られてくる発表論文情報をまとめ、著者氏名毎に論文の題名を整理した一覧表(以降、一覧表Aと呼ぶ)を作成したいとする。共著者が存在する場合には、著者の数だけ同じ論文の題名を重複して一覧表Aに掲載する。具体例を図8に示す。図8において、入力文書801、802は送られてくる発表論文情報であり、出力文書804は作成したい一覧表Aである。

#### 【0010】

このとき、前記発表論文情報は、著者氏名や論文の題名の他に、投稿先学会名や論文概要などの情報を含む、S G M L 形式で記述された構造化文書であるとする。すなわち、前記発表論文情報は、

1. <学会名>および</学会名>という

タグで囲まれた投稿先学会名を表す部分構造

2. <題名>および</題名>というタグで

囲まれた論文の題名を表す部分構造

3. <氏名>および</氏名>というタグで

囲まれた著者氏名を表す部分構造の繰り返し

4. <概要>および</概要>というタグで

囲まれた論文概要を表す部分構造

をこの順序で並べて、全体を<論文情報>および</論文情報>というタグで囲んだ形式で表現する。ただし、著者氏名を表す部分構造は、共著者が存在する場合に繰り返される。例えば、図8の入力文書801、802をS G M L 形式で記述した例を図9に示す。図9において、符号901は入力文書801をS G M L 形式で記述した構造化文書、符号902は入力文書802をS G M L 形式で記述した構造化文書である。

#### 【0011】

また、前記作成したい一覧表Aは、

1. <氏名>および</氏名>というタグで

囲まれた著者氏名を表す部分構造

2. <題名>および</題名>というタグで

囲まれた論文の題名を表す部分構造

をこの順序で並べて全体を<表要素>および</表要素>というタグで囲んだ部分構造を、さらに複数並べて、全体を<一覧表A>および</一覧表A>というタグで囲んだ形式で表現する。例えば、図8の出力文書804をS G M L 形式で記述した例を図10に示す。

#### 【0012】

以上の前提のもとで、前記特開平6-259421号公報に代表されるパター

ン記述を用いて、前記一覧表Aを作成する処理例を説明する。

### 【0013】

まず、従来の方式で扱うことのできる例として、入力となる前記発表論文情報には共著者がないと仮定する。すなわち、前記発表論文情報は、

1. <学会名>および</学会名>というタグで

囲まれた投稿先学会名を表す部分構造

2. <題名>および</題名>というタグで

囲まれた論文の題名を表す部分構造

3. <氏名>および</氏名>というタグで

囲まれた著者氏名を表す部分構造

4. <概要>および</概要>というタグで

囲まれた論文概要を表す部分構造

をこの順序で並べて、全体を<論文情報>および</論文情報>というタグで囲んだ形式で表現するとする。

### 【0014】

1つのパターンによって抽出できる部分構造が1つに限られる従来の方式では、抽出したい部分構造毎に用意したパターン記述を用いて、前記発表論文情報から部分構造を抽出する。すなわち、

- a. 前記論文の題名を抽出するパターンを用いて、

前記発表論文情報から、前記論文の題名を表す

部分構造を抽出し、

- b. 前記著者氏名を抽出するパターンを用いて、

前記発表論文情報から、前記著者氏名を表す

部分構造を抽出する。

そして、抽出した各部分構造の組を作り、前記一覧表Aにまとめあげる。すなわち、

- c. 前記a. で抽出した部分構造を、抽出した順に

1つ取り出し、前記b. で抽出した部分構造を、

抽出した順に1つ取り出して、前記2つの部分

構造を並べて<表要素>および</表要素>というタグで囲む、という処理を、前記 a. および b. で抽出した全ての部分構造に対して順に行う。

#### 【0015】

d. 前記 c. の処理によって得られた部分構造を並べて、全体を<一覧表 A>および</一覧表 A>というタグで囲む。

以上の処理により、共著者が不在の発表論文情報から一覧表 A を作成することができる。

#### 【0016】

次に、入力となる前記発表論文情報には共著者が不在という仮定をなくし、共著者の存在する発表論文情報が入力文書に含まれる場合を考える。入力文書に、共著者の存在する発表論文情報が含まれている場合には、前記 c. の処理は、c'. 同じ発表論文情報の中にあつて、論文の題名、著者氏名の順で並んでいる全ての組合せを求め、それぞれを、<表要素>および</表要素>というタグで囲む。

としなければならない。しかしながら、従来の方式では、論文の題名と著者氏名は、それぞれ別々のパターンで抽出するため、前記 c'. の処理で必要となる「同じ発表論文情報の中に並んでいる」という関係を規定することができない。例えば、図 9 の構造化文書 901 および 902 から、図 10 のような出力文書を作成したい場合でも、図 11 に示すように、構造化文書 901 および 902 から抽出した論文の題名を表す部分構造 1102 と著者氏名を表す部分構造 1103 とをどのように組み合わせたらよいか分からないため、出力文書 1101 を作成することができない。従って、従来の方式では、上記例で、共著者の存在する発表論文情報から一覧表を作成する場合などのように、抽出した部分構造間の関係が指定される文書編集処理はできないという問題点があった。

#### 【0017】

よって、抽出箇所間のつながりを維持して部分構造を抽出することのできる編

集箇所記述要素、および前記編集箇所記述要素を解釈して部分構造の抽出処理を実行する手段が必要となった。

#### 【 0 0 1 8 】

(2-2) 特開平6-259421号公報に代表される従来の文書編集方式のパターン記述において、複数の抽出箇所から部分構造を同時に1つのかたまりとして抽出するパターンが記述できるとした場合でも、抽出した部分構造のかたまりを利用する際に、部分構造のかたまりから個々の部分構造を取り出す処理、あるいは、部分構造のかたまりの中の部分構造の並んでいる順番を変えるための処理などが別途必要となるという問題点があった。

#### 【 0 0 1 9 】

例を用いて前記(2-2)の問題点を説明する。例えば、図9の構造化文書901および902から前記方式により一覧表Aを作成した例を図12に示す。図12において、構造化文書901および902から、題名、氏名の並びからなる部分文書のかたまりを抽出することができたとする。このとき、前記抽出した部分文書1202をそのまま用いて一覧表Aを作成した場合には、図12の出力文書1201のように、著者氏名、論文の題名の順に並ぶべきなのに対し、論文の題名、著者氏名の順になってしまう。

#### 【 0 0 2 0 】

よって、抽出箇所間のつながりを維持して部分構造を抽出し、かつ、パターン上の抽出箇所毎に別々に扱うことのできる編集箇所記述要素、および前記編集箇所記述要素を解釈して部分構造の抽出処理を実行する手段が必要となった。

#### 【 0 0 2 1 】

(2-3) 特開平6-259421号公報に代表される従来の文書編集方式のパターン記述の構文要素には、順序関係を表す記号はあるが、順不同の並列関係を表す記号はない。特にSGML形式で記述された構造化文書を扱う場合などは、複数種類の部分構造を任意の順序で並べて記述することも可能であるため、前記任意の順序で並べて記述された部分構造とマッチするパターンを、順序関係を表す記号を用いて記述すると、記述が繁雑になるという問題点があった。

#### 【 0 0 2 2 】

例を用いて前記（２－３）の問題点を説明する。例えば、前記発表論文情報の表現形式のうち、投稿先学会名を表す部分構造と、論文の題名を表す部分構造との記述順序は問わない記述形式の発表論文情報（以降、発表論文情報２と呼ぶ）から、投稿先学会名と論文の題名との対応を整理した一覧表（以降、一覧表Ｂと呼ぶ）を作成したいとする。

#### 【００２３】

すなわち、

- １． 次の２種類の部分構造を、任意の順序で

両方とも記述

- １－１． <学会名>および</学会名>という

タグで囲まれた投稿先学会名を表す部分構造

- １－２． <題名>および</題名>というタグ

で囲まれた論文の題名を表す部分構造

- ２． <氏名>および</氏名>というタグで

囲まれた著者氏名を表す部分構造の繰り返し

- ３． <概要>および</概要>というタグで

囲まれた論文概要を表す部分構造

をこの順序で並べ、全体を<論文情報２>および</論文情報２>というタグで囲んだ形式で表現したものが発表論文情報２であり、また、

- １． <学会名>および</学会名>という

タグで囲まれた投稿先学会名を表す部分構造

- ２． <題名>および</題名>というタグで

囲まれた論文の題名を表す部分構造

をこの順序で並べて全体を<表要素>および</表要素>というタグで囲んだ部分構造を、さらに複数並べて、全体を<一覧表Ｂ>および</一覧表Ｂ>というタグで囲んだ形式で表現したものが作成したい一覧表Ｂであるとする。

#### 【００２４】

論文情報２のＳＧＭＬ形式による記述例を図２７の入力文書２７０１および入力文書２７０２に、また、一覧表ＢのＳＧＭＬ形式による記述例を図２７の出力文書２７０３に示

す。

### 【0025】

以上の前提のもとで、前記特開平6-259421号公報に代表されるパターン記述を用いて、前記発表論文情報2を表す入力文書2601と入力文書2602から前記一覧表Bを表す出力文書2604を文書編集方式2603により作成することを考える(図26参照)。

### 【0026】

前記投稿先学会名を表す部分構造および論文の題名を表す部分構造を抽出するためのパターン記述は、

1. 投稿先学会名を表す部分構造、論文の題名を表す部分構造の順で記述された場合
2. 論文の題名を表す部分構造、投稿先学会名を表す部分構造の順で記述された場合

の2通りを考慮する必要があるため、前記特開平6-259421号公報では、

1. 投稿先学会名を表す部分構造、論文の題名を表す部分構造の順で記述された場合の  
パターン記述

論文情報／(学会名#題名)

2. 論文の題名を表す部分構造、投稿先学会名を表す部分構造の順で記述された場合の  
パターン記述

論文情報／(題名#学会名)

の2通りのパターンを記述する必要がある。

### 【0027】

前記例では、記述順序を問わない部分構造は2つであるためパターンを2通り用意すれば十分であったが、一般に、記述順序を問わない部分構造がn個あった場合には、パターンをn!通り用意する必要がある(!は階乗計算)。例えば、記述順序を問わない部分構造が5個あった場合には、120通りのパターンを用意しなければならず、前記特開平6-259421号公報に代表される従来の文書

編集方式では、パターン記述が非常に繁雑となる。

#### 【 0 0 2 8 】

よって、順不同の並列関係を記述することのできる編集箇所記述要素、および前記編集箇所記述要素を解釈して部分構造の抽出処理を実行する手段が必要となった。

#### 【 0 0 2 9 】

##### 【課題を解決するための手段】

本発明の文書編集方式は、  
複数の構造化文書を読み込み、  
文書の構造を単位として複数の抽出箇所を指定するパターン記述に従って、抽出箇所間のつながりを維持しながら、前記構造化文書から部分構造を抽出するようにしている。

#### 【 0 0 3 0 】

本発明の文書編集方式は、  
1つ以上の構造化文書を順次読み込む構造化文書入力手段と、  
前記構造化文書入力手段の読み込んだ入力文書の一部である部分構造を指定する、文書タグ、文書文字パターン、ワイルドカードタグ、否定指定子、引用指定子、順序接続子、階層接続子、括弧によって構成される編集箇所記述要素を、AND接続子、OR接続子、括弧で結合した編集箇所記述が1つ以上記述されている文書処理記述手段と、  
前記部分構造を、前記文書処理記述手段に記述された編集箇所記述中の引用指定子毎に分け、かつ、前記入力文書と前記編集箇所記述とのマッチが成立した順に蓄積する文書蓄積手段と、  
前記文書処理記述手段から編集箇所記述を取り出し、前記編集箇所記述を構成する各編集箇所記述要素と入力文書とのマッチを試み、マッチが成立した前記編集箇所記述中に記述された引用指定子に対応する部分構造を前記文書蓄積手段に蓄積し、また、前記編集箇所記述中のAND接続子で結合された編集箇所記述要素の少なくとも1つのマッチが成立しない場合には、前記AND接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除し、また

、前記編集箇所記述中のOR接続子で結合された編集箇所記述要素のどのマッチも成立しない場合には、前記OR接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除して抽出処理を行う文書編集処理実行手段と、

前記文書蓄積手段に蓄積された文書をそのまま、あるいは加工して出力する文書出力手段と

から構成されるようにしてもよい。

### 【0031】

#### [作用]

本発明の文書編集方式は、文書の構造を単位として編集箇所を指定する編集箇所記述を解釈し、部分構造を抽出する文書編集処理実行手段を備える。これにより、文書の構造を単位としてパターン記述し、また、文書編集処理を行うことが可能となり、前記「発明が解決しようとする課題」の（１）の問題点が解決する。

### 【0032】

また、本発明の文書編集方式は、編集箇所記述中に複数の引用指定子（抽出箇所を指定する記号）が記述でき、かつ、文書編集処理実行手段のマッチ処理によって抽出された部分構造は、マッチが成立した順に、引用指定子毎に分けて文書蓄積手段に蓄積されている。これにより、１つの編集箇所記述によって複数の部分構造を同時に抽出し、かつ、抽出した部分構造を、抽出箇所間のつながりを維持しつつ、編集箇所記述中の引用指定子毎に別々に扱うことができ、前記「発明が解決しようとする課題」の（２－１）および（２－２）の問題点が解決する。

### 【0033】

また、本発明の文書編集方式は、編集箇所記述中に順不同の並列関係を表すAND接続子が記述でき、前記編集箇所記述を解釈し、部分構造を抽出する文書編集処理実行手段を備える。これにより、SGML形式など、複数種類の部分構造を任意の順序で並べて記述することが可能な形式で記述された構造化文書を扱う場合でも、記述が簡潔となり、前記「発明が解決しようとする課題」の（２－３）の問題点が解決する。

#### 【 0 0 3 4 】

##### 【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して詳細に説明する。

#### 【 0 0 3 5 】

以降では、入力文書、出力文書とも、S G M L形式である例について説明する。

#### 【 0 0 3 6 】

まず、本発明における編集箇所記述の各構成要素の定義と、表記形式の一例を説明する。なお、編集箇所記述の各構成要素は、一般に、構造化文書中の複数の部分構造とマッチする。どの部分構造を選択してマッチ処理を進めるかは、マッチ処理を実行する文書編集処理実行手段の処理方式に依存する（例えば、任意の部分構造を始点として編集箇所記述要素とマッチする全ての部分構造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする最初の部分構造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする最後の部分構造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も外側の部分構造を選択する方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も内側の部分構造を選択する方式などの処理方式がある）。

#### 【 0 0 3 7 】

（a）文書タグ：部分構造に付けられた名前を<と>とで囲んで表記する

マッチさせたい部分構造を特定する名前を文書タグとして指定することにより、名前の合致する部分構造とマッチする。例えば、S G M L形式では、部分構造に付けられた名前として、タグ名を利用することが考えられる。

#### 【 0 0 3 8 】

例として、図13を用いて説明する。符号1301は編集箇所記述、符号1302はS G M L形式の構造化文書である。図13の例では、編集箇所記述1301中には、文書タグ<注意事項>が記述されている。編集箇所記述1301と構造化文書1302とのマッチ処理の過程において、前記文書タグは、図13に示す通り、構造化文書1302の現

在着目している部分構造中に並んでいる部分構造のうち、” 注意事項” という名前のタグによって囲まれた部分構造とマッチする。

#### 【0039】

(b) 文書文字パターン：正規表現記号を含む文字列からなる文字パターンを” と” とで囲んで表記する

文字列、もしくは、文字列に加えて `sed` などのUNIXコマンドで一般的な正規表現記号を文書文字パターンとして指定することにより、合致する文字パターンの出現する文書データとマッチする。なお、文書データとは、例えば、SGML形式の構造化文書では、タグを除く文章部分のことである。

#### 【0040】

例として、図14を用いて説明する。図14の符号1401は編集箇所記述、符号1402はSGML形式の構造化文書である。図14の例では、編集箇所記述1401中には、文書文字パターン” the” が記述されている。編集箇所記述1401と構造化文書1402とのマッチ処理の過程において、前記文書文字パターンは、図14に示す通り、構造化文書1402の現在着目している部分構造中に並んでいる文書データのうち、” the” という文字列を含んだ文書データとマッチする。

#### 【0041】

(c) ワイルドカードタグ：<\*>と表記する

後述する階層関係にある任意個の任意の名前の文書タグとマッチする。

#### 【0042】

例として、図15を用いて説明する。図15の符号1501は編集箇所記述、符号1502はSGML形式の構造化文書である。図15の例では、編集箇所記述1501中には、ワイルドカードタグ<\*>が記述されており、また、構造化文書1502は、” 本文” の中に” 章” があり、” 章” の中に” 節” があり、” 節” の中に” 項” があり、” 項” の中は文書データのみ、という文書構造である。編集箇所記述1501と構造化文書1502とのマッチ処理の過程において、図15に示す通り、構造化文書1502の現在着目している部分構造が” 章” である場合には、前記ワイルドカードタグは

#### 1. 「空」の構造とマッチ（0個の階層とマッチ）

2. ” 節” の構造とマッチ

3. ” 節” の中に” 項” がある構造とマッチ

の3通りの部分構造とマッチする。

#### 【0043】

(d) 否定指定子：！で表記する

否定指定子の直後の記述とのマッチが成り立たない部分構造とマッチする。

#### 【0044】

例として、図16を用いて説明する。図16の符号1601は編集箇所記述、符号1602はSGML形式の構造化文書である。図16の例では、編集箇所記述1601中には、否定指定子！の直後に文書文字パターン” this” が記述されている。編集箇所記述1601と構造化文書1602とのマッチ処理の過程において、前記否定指定子および前記否定指定子の直後の文書文字パターンは、図16に示す通り、構造化文書1602の現在着目している部分構造中に並んでいる文書データのうち、” this” という文字列を含まない文書データとマッチする。

#### 【0045】

(e) 引用指定子：%で表記する

引用指定子の直後の記述とマッチした部分構造を抽出する。

#### 【0046】

例として、図17を用いて説明する。図17の符号1701は編集箇所記述、符号1702はSGML形式の構造化文書である。図17の例では、編集箇所記述1701中には、引用指定子%の直後に文書タグ<注意事項>が記述されている。編集箇所記述1701と構造化文書1702とのマッチ処理の過程において、前記文書タグは、図17に示す通り、構造化文書1702の現在着目している部分構造中に並んでいる部分構造のうち、” 注意事項” という名前のタグによって囲まれた部分構造とマッチするため、編集箇所記述1701によって、前記マッチした部分構造が抽出される。

#### 【0047】

(f) 順序接続子：，で表記する

順序接続子の直前の記述にマッチする部分構造（部分構造Aとする）と、前記順序接続子の直後の記述にマッチする部分構造（部分構造Bとする）とが、同じ

部分構造の中に部分構造 A、部分構造 B の順序で並んでいる場合にマッチが成立する。部分構造 A と部分構造 B との間に他の部分構造が存在していてもマッチが成立する。

#### 【 0 0 4 8 】

例として、図18を用いて説明する。図18の符号1801は編集箇所記述、符号1802は S G M L 形式の構造化文書である。図18の例では、編集箇所記述1801中には、順序接続子、の直前に文書タグ<はじめに>が記述されており、かつ、前記順序接続子の直後に文書タグ<おわりに>が記述されている。編集箇所記述1801と構造化文書1802とのマッチ処理の過程において、図18に示す通り、順序接続子の直前の文書タグと、構造化文書1802の現在着目している部分構造中に並んでいる”はじめに” という名前のタグによって囲まれた部分構造とがマッチし、さらに、前記順序接続子の直後の文書タグと、前記マッチした部分構造より後方に並んでいる” おわりに” という名前のタグによって囲まれた部分構造とがマッチするため、前記順序接続子と、前記順序接続子の前後の文書タグとで構成された記述のマッチが成立する。

#### 【 0 0 4 9 】

(g) 階層接続子：2つの記述の間に何も文字を入れないことで表記する

階層接続子の直前の記述にマッチする部分構造（部分構造 C とする）の中に、前記階層接続子の直後の記述にマッチする部分構造（部分構造 D とする）がある場合にのみマッチが成立する。なお、部分構造 C の中で、部分構造 D の前に他の部分構造が並んでいてもマッチは成立する。

#### 【 0 0 5 0 】

例として、図19を用いて説明する。図19の符号1901は編集箇所記述、符号1902は S G M L 形式の構造化文書である。図19の例では、編集箇所記述1901中には、階層接続子の直前に文書タグ<節>が記述されており、かつ、前記並列接続子の直後に文書タグ<項>が記述されている。また、構造化文書1902は、” 本文” の中に” 章” があり、” 章” の中に” 節” があり、” 節” の中に” 項” があり、” 項” の中は文書データのみ、という文書構造である。編集箇所記述1901と構造化文書1902とのマッチ処理の過程において、図19に示す通り、階層接続子の直前の

文書タグと、構造化文書1902の現在着目している部分構造中に並んでいる” 節” という名前のタグによって囲まれた部分構造とがマッチし、さらに、前記階層接続子の直後の文書タグと、前記マッチした部分構造の中に並んでいる” 項” という名前のタグによって囲まれた部分構造とがマッチするため、前記階層接続子と、前記階層接続子の前後の文書タグとで構成された記述のマッチが成立する。

#### 【0051】

(h) 括弧：（および）で表記する

括弧によって囲まれた記述のマッチが優先処理される。

#### 【0052】

(i) AND接続子：&で表記する

AND接続子の直前の記述にマッチする部分構造（部分構造Eとする）と、前記AND接続子の直後の記述にマッチする部分構造（部分構造Fとする）とが、同じ部分構造中に並んでいる場合にマッチが成立する。部分構造Eと部分構造Fとが並んでいる順序は問わず、また、2つの部分構造の間に他の部分構造が存在していてもマッチが成立する。

#### 【0053】

例として、図20を用いて説明する。図20の符号2001は編集箇所記述、符号2002はSGML形式の構造化文書である。図20の例では、編集箇所記述2001中には、AND接続子&の直前に文書タグ<おわりに>が記述されており、かつ、前記AND接続子の直後に文書タグ<はじめに>が記述されている。編集箇所記述2001と構造化文書2002とのマッチ処理の過程において、図20に示す通り、AND接続子の直前の文書タグと、構造化文書2002の現在着目している部分構造中に並んでいる” おわりに” という名前のタグによって囲まれた部分構造とがマッチし、さらに、AND接続子の直後の文書タグと、前記現在着目している部分構造中に並んでいる” はじめに” という名前のタグによって囲まれた部分構造とがマッチするため、前記AND接続子と、前記AND接続子の前後の文書タグとで構成された記述のマッチが成立する。

#### 【0054】

(j) OR接続子：|で表記する

OR 接続子の直前の記述にマッチする部分構造（部分構造Gとする）と、前記OR 接続子の直後の記述にマッチする部分構造（部分構造Hとする）の、少なくとも一方の部分構造が存在する場合にマッチが成立する。

#### 【0055】

例として、図21を用いて説明する。図21の符号2101は編集箇所記述、符号2102はSGML形式の構造化文書である。図21の例では、編集箇所記述2101中には、OR 接続子 | の直前に文書タグ<おわりに>が記述されており、かつ、前記OR 接続子の直後に文書タグ<はじめに>が記述されている。編集箇所記述2101と構造化文書2102とのマッチ処理の過程において、図21に示す通り、OR 接続子の直前の文書タグとマッチする部分構造が、構造化文書2102の現在着目している部分構造中に並んでいるかを調べ、さらに、OR 接続子の直後の文書タグとマッチする部分構造が、構造化文書2102の現在着目している部分構造中に並んでいるかを調べる。OR 接続子の直後の文書タグとマッチする部分構造が存在するため、前記OR 接続子と、前記OR 接続子の前後の文書タグとで構成された記述のマッチが成立する。

#### 【0056】

図1は本発明の第1の実施の形態の構成図である。図1を参照すると、本実施の形態の文書編集方式は、  
構造化文書を順次読み込む構造化文書入力手段101と、  
前記構造化文書入力手段の読み込んだ入力文書の部分構造を指定する編集箇所記述が1つ以上記述されている文書処理記述手段102と、  
前記部分構造を、前記文書処理記述手段に記述された編集箇所記述中の引用指定子毎に分け、かつ、前記入力文書と前記編集箇所記述とのマッチが成立した順に蓄積する文書蓄積手段103と、  
前記文書処理記述手段から編集箇所記述を取り出し、前記編集箇所記述を構成する各編集箇所記述要素と入力文書とのマッチを試み、マッチが成立した前記編集箇所記述中に記述された引用指定子に対応する部分構造を前記文書蓄積手段に蓄積し、また、前記編集箇所記述中のAND 接続子で結合された編集箇所記述要素の少なくとも1つのマッチが成立しない場合には、前記AND 接続子で結合され

た各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除し、また、前記編集箇所記述中のOR接続子で結合された編集箇所記述要素のどのマッチも成立しない場合には、前記OR接続子で結合された各編集箇所記述要素に対応する部分構造を前記文書蓄積手段から削除して抽出処理を行う文書編集処理実行手段104と、

前記文書蓄積手段に蓄積された文書をそのまま、あるいは加工して出力する文書出力手段105とを有する。

#### 【0057】

次に、動作について図22、図23、図24、図25を参照して説明する。

#### 【0058】

図22は、第1の実施の形態の全体処理動作を示す流れ図である。

#### 【0059】

文書編集の対象となる複数の構造化文書が与えられると、構造化文書入力手段101は前記複数の構造化文書を1つ読み込む(ステップ2201)。文書編集処理実行手段104は、文書処理記述手段102から編集箇所記述を1つ取り出して(ステップ2202)抽出処理を実行する(ステップ2203)。抽出処理2203の実行が終了すると、次に処理すべき編集箇所記述がある場合(ステップ2204)には、編集箇所記述を1つ取り出す処理に戻り、また、処理すべき編集箇所記述がない場合には、次に処理すべき構造化文書があるかを調べる(ステップ2205)。全ての構造化文書に対する処理がまだ完了していない場合には、ステップ2201の文書を1つ読み込む処理に戻り、また、全ての構造化文書に対する処理が完了した場合には、文書出力手段105が、文書蓄積手段103に蓄積された部分構造をそのまま、あるいは加工して出力する(ステップ2206)。

#### 【0060】

前記抽出処理における動作フローを、図23を用いて説明する。

#### 【0061】

抽出処理では、文書編集処理実行手段104が、OR接続子を区切りとして編集箇所記述を複数の記述に分解し(ステップ2301)、前記分解した編集箇所記述の1つを選び(ステップ2302)AND接続子の処理を実行する(ステップ2303)。前記ス

テップ2303のAND接続子の処理の実行が終ると、文書編集処理実行手段104は、前記分解した編集箇所記述の中で未処理のものがある場合(ステップ2304)には、ステップ2302の分解した編集箇所記述の1つを選ぶ処理に戻り、また、前記分解した編集箇所記述全てに対する処理が完了した場合には、前記分解した各記述のマッチの状況を調べて、分解したどの記述もマッチが成立していない場合(ステップ2305)には、対応する部分構造を文書蓄積手段103から削除する(ステップ2306)。

#### 【0062】

前記AND接続子の処理における動作フローを、図24を用いて説明する。

#### 【0063】

AND接続子の処理では、文書編集処理実行手段104が、AND接続子を区切りとして編集箇所記述を複数の記述に分解し(ステップ2401)、前記分解した編集箇所記述の1つを選び(ステップ2402)マッチ処理を実行する(ステップ2403)。ステップ2403のマッチ処理実行が終ると、文書編集処理実行手段104は、前記分解した編集箇所記述の中で未処理のものがある場合(ステップ2404)には、ステップ2402の分解した編集箇所記述の1つを選ぶ処理に戻り、また、前記分解した記述全てに対する処理が完了した場合には、前記分解した各記述のマッチの状況を調べて、マッチが不成立の記述が1つでもある場合(ステップ2405)には、対応する部分構造を文書蓄積手段103から削除する(ステップ2406)。

#### 【0064】

前記マッチ処理における動作フローを、図25を用いて説明する。

マッチ処理では、文書編集処理実行手段104が、処理対象となる編集箇所記述の先頭要素を着目点とし(ステップ2501)、前記編集箇所記述の着目点における要素が括弧で囲まれている場合(ステップ2502)には、前後の括弧をはずして(ステップ2503)抽出処理を実行し(ステップ2504)、また、前記編集箇所記述の着目点における要素が括弧で囲まれていない場合には、前記要素と、入力文書中の現在着目している部分構造とのマッチを試みる(ステップ2505)。ステップ2504の抽出処理もしくはステップ2505のマッチ完了後、文書編集処理実行手段104は、前記編集箇所記述の着目点が記述の末尾に達していない場合(ステップ2506)には、前記

編集箇所記述の着目点の要素と前記着目点の次の要素との間にある接続子に応じて入力文書中の現在着目している部分構造を変え(ステップ2507)、編集箇所記述の着目点を次の要素に移して(ステップ2508)、ステップ2502の編集箇所記述の着目点の要素が括弧で囲まれているか判定する処理へ戻る。また、前記編集箇所記述の着目点が記述の末尾に達した場合には、前記編集箇所記述と入力文書とのマッチの結果を調べてマッチが成立した場合には(ステップ2509)、さらに、前記編集箇所記を調べて引用指定子があった場合には(ステップ2510)、前記引用指定子に対応する部分構造を前記入力文書から抽出し、引用指定子毎に分けて文書蓄積手段103に蓄積する(ステップ2511)。

#### 【0065】

##### 【実施例】

次に、本発明の第1の実施の形態の実施例を3つの具体例を用いて動作を詳細に説明する。

#### 【0066】

まず、実施例1を、図8の、発表論文情報から文書編集方式803により論文の題名と著者氏名を抽出して一覧表Aを作成する例を用いて説明し、発明が解決しようとする課題の(2-1)および(2-2)が解決することを示す。

#### 【0067】

入力文書として、図9の構造化文書901および構造化文書902の2つが与えられている例を説明する。

#### 【0068】

また、発表論文情報から一覧表Aの作成に必要な部分構造を抽出するための編集箇所記述として、

%<題名>、%<氏名>

が文書処理記述手段に記述されているとする。なお、以降では、前記編集箇所記述を編集箇所記述Aと呼ぶことにする。

#### 【0069】

また、文書出力手段による加工方法は、次の通りであるとする。なお、以降では、下記加工方法を、加工方法Aと呼ぶことにする。

### 【0070】

1. 編集箇所記述Aの2つめの引用指定子と

対応する部分構造

2. 編集箇所記述Aの1つめの引用指定子と

対応する部分構造

を文書蓄積手段からそれぞれ1つずつ抽出して

前記順番に並べたものを<表要素>というタグで

囲んで表要素を作り、文書蓄積手段中の全て

の部分構造に対して前記処理を実行して得られた

表要素を並べて、全体を<一覧表A>というタグで

囲んで出力する。

以上の条件の下での動作を説明する。

構造化文書入力手段101が、1つめの入力文書として、構造化文書901を読み込む(ステップ2201)。文書編集処理実行手段104は、文書処理記述手段102から前記編集箇所記述Aを取り出して(ステップ2202)、構造化文書901に対する抽出処理を実行する(ステップ2203)。

### 【0071】

抽出処理では、OR接続子を区切りとして、編集箇所記述Aを複数の記述に分解する(ステップ2301)。編集箇所記述A中にはOR接続子はないため、分解された記述は、編集箇所記述A全体からなるもの1つだけとなる。前記編集箇所記述Aを処理対象として(ステップ2302)、AND接続子の処理を実行する(ステップ2303)。

### 【0072】

AND接続子の処理では、AND接続子を区切りとして編集箇所記述Aを複数の記述に分解する(ステップ2401)。編集箇所記述A中にはAND接続子はないため、分解された記述は、編集箇所記述A全体からなるもの1つだけとなる。前記編集箇所記述Aを処理対象として(ステップ2402)、マッチ処理を実行する(ステップ2403)。

### 【0073】

マッチ処理では、まず、編集箇所記述 A の先頭要素である%<題名>を着目点とする(ステップ2501)。前記着目点の要素と入力文書とのマッチの実行方式には何通りかあるが、ここでは、文書の任意の部分構造を始点として編集箇所記述要素とマッチする全ての部分構造を求める方式を用いて、

1. 編集箇所記述の先頭要素に限っては、入力文書

中の全ての部分構造を対象としてマッチを開始し、

かつ、

2. マッチする全ての部分構造を求める

方式とする。前記マッチの実行方式の 1. より、前記着目点の要素%<題名>と、構造化文書901中の全ての部分構造とのマッチを試み、

<題名>

××に関する研究

</題名>

とマッチする(ステップ2505)。また、前記着目点の要素と前記着目点の次の要素との間には順序接続子があるため、前記マッチした部分構造の後に並んでいる構造が、構造化文書901の現在着目している部分構造となる(ステップ2507)。

#### 【 0 0 7 4 】

前記編集箇所記述 A の着目点が次の要素%<氏名>に移され(ステップ2508)、前記構造化文書901中の現在着目している部分構造とのマッチが試みられる。前記マッチの実行方式の 2. に従ってマッチする全ての部分構造を求めるため、前記着目点の要素%<氏名>は、

<氏名>

○× 太郎

</氏名>

<氏名>

×○ 花子

</氏名>

の 2 通りの部分構造とマッチする(ステップ2505)。

### 【 0 0 7 5 】

着目点は前記編集箇所記述Aの末尾に達したため(ステップ2506)、編集箇所記述Aの各要素と構造化文書901とのマッチを終え、各引用指定子に対応する部分構造を抽出する。この際、抽出箇所間のつながりを維持するため、例えば、編集箇所記述Aの要素%<題名>と構造化文書901とマッチした部分構造と、編集箇所記述Aの要素%<氏名>と構造化文書901とマッチした2つの部分構造との組合せを求めて、文書蓄積手段103に蓄積する(ステップ2511)。すなわち、例えば、次のような情報を文書蓄積手段103に蓄積してマッチ処理を終了する。

### 【 0 0 7 6 】

編集箇所記述Aの1つめの引用指定子に対応する

部分構造

< 題名 >

××に関する研究

< / 題名 >

< 題名 >

××に関する研究

< / 題名 >

編集箇所記述Aの2つめの引用指定子に対応する

部分構造

< 氏名 >

○× 太郎

< / 氏名 >

< 氏名 >

×○ 花子

< / 氏名 >

編集箇所記述Aを処理対象としたマッチ処理が完了すると、AND接続子の処理に戻る。AND接続子の処理では、分解した記述に対する処理は全て完了した

ため(ステップ2404)、前記編集箇所記述Aと入力文書とのマッチの状況を調べる(ステップ2405)。前記編集箇所記述Aはマッチが成立したため、文書蓄積手段の削除は行わず、AND接続子の処理を終了する。

#### 【0077】

AND接続子の処理が完了すると、抽出処理に戻る。抽出処理では、分解した記述に対する処理は全て完了したため(ステップ2304)、前記編集箇所記述Aと入力文書とのマッチの状況を調べる(ステップ2305)。前記編集箇所記述Aはマッチが成立したため、文書蓄積手段の削除は行わず、抽出処理を終了する。

#### 【0078】

抽出処理を完了すると、全体の処理に戻る。全体の処理では、構造化文書901に対して処理すべき編集箇所記述はもうないため(ステップ2204)、構造化文書入力手段101が、2つめの入力文書として、構造化文書902を読み込む(ステップ2201)。文書編集処理実行手段104は、文書処理記述手段102から前記編集箇所記述Aを取り出して(ステップ2202)、構造化文書902に対する抽出処理を実行する(ステップ2203)。

#### 【0079】

構造化文書902に対する抽出処理は、前記構造化文書901に対する抽出処理とほぼ同じであるため、詳細な説明は省き、結果だけを示す。処理の結果、前記編集箇所記述Aの1つめの引用指定子に対応する部分構造として、構造化文書902から、

<題名>

〇〇に関する研究

</題名>

が抽出され、また、前記編集箇所記述Aの2つめの引用指定子に対応する部分構造として、構造化文書902から、

<氏名>

〇〇 次郎

</氏名>

が抽出され、文書蓄積手段103に蓄積される。

### 【0080】

構造化文書902に対する抽出処理が終了すると、全体の処理に戻る。全体の処理では、構造化文書902に対して処理すべき編集箇所記述はもうなく(ステップ2204)、また、全ての構造化文書に対しての抽出処理が完了したため(ステップ2205)、文書出力手段105が、文書蓄積手段103に蓄積された部分構造を加工する(ステップ2206)。

### 【0081】

この時点で文書蓄積手段103に蓄積されている部分構造を、文書蓄積手段103の格納形式の一例を用いて以下に示す。

### 【0082】

編集箇所記述Aの1つめの引用指定子に対応する

部分構造

<題名>

××に関する研究

</題名>

<題名>

××に関する研究

</題名>

<題名>

〇〇に関する研究

</題名>

編集箇所記述Aの2つめの引用指定子に対応する

部分構造

<氏名>

〇× 太郎

</氏名>

<氏名>

×○ 花子

</氏名>

<氏名>

○○ 次郎

</氏名>

このように、文書蓄積手段103には、引用指定子毎に対応する部分構造が、部分構造の並びとして蓄積されている。しかも、それぞれの部分構造の並びの先頭の部分構造は、最初にマッチが成立した際に抽出したものであり、それぞれの部分構造の並びの2番目の部分構造は、2番目にマッチが成立した際に抽出したものである、といったように、並びの順番という形で、各引用指定子に対応する部分構造の間のつながりを維持している。

#### 【0083】

文書出力手段105は、これらの部分構造から、前記加工方法Aにより、図10のような一覧表Aを作成し、出力する(ステップ2206)。

#### 【0084】

以上のように、本発明では、例えば前記編集箇所記述Aのように、1つのパターンに複数の抽出箇所を指定することができ、しかも、例えば、文書蓄積手段103に蓄積された部分構造の並びの順番という形で、各引用指定子に対応する部分構造の間のつながりを維持して部分構造を抽出することができ、さらに、前記加工方法Aのように、構造を分解する処理と特別に用意することなく、抽出した部分構造を、引用指定子毎に別々に扱うことができる。

#### 【0085】

なお、前記マッチ処理におけるマッチの実行方式として文書の任意の部分構造を始点として編集箇所記述要素とマッチする全ての部分構造を求める方式を用いたが、用途や処理速度、使用する作業領域の大きさなどに応じて、任意の部分構造を始点として編集箇所記述要素とマッチする最初部分構造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする最後の部分構

造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も外側の部分構造を選択する方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も内側の部分構造を選択する方式もしくは、最も外側の構造を始点として前記編集箇所記述要素とマッチする全ての部分構造を求める方式のいずれの方式、もしくはその他の方式を用いても良い。例えば、任意の部分構造を始点として編集箇所記述要素とマッチする最初の部分構造を求める方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする最後の部分構造を求める方式は、マッチする組合せの1つを求めれば十分である場合などに適用することができ、任意の部分構造を始点として編集箇所記述要素とマッチする全ての部分構造を求める方式などと比較して処理速度が早く、使用する作業領域の大きさも小さいという利点がある。また、例えば、入力文書中に、

```
<itemize>
  <item> 選択枝1  </item>
  <itemize>
    <item> 副選択枝A  </item>
    <item> 副選択枝B  </item>
  </itemize>
  <item> 選択枝2  </item>
  <item> 選択枝3  </item>
</itemize>
```

など、<itemize>の内側にさらに<itemize>が含まれているといったように同種の構造がネストしている場合において、一番外側の<itemize>とマッチさせたい場合、あるいは、一番内側の<itemize>とマッチさせたい場合には、前述した、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も外側の部分構造を選択する方式もしくは、任意の部分構造を始点として編集箇所記述要素とマッチする部分構造のうち、包含関係にある部分構造は最も内側の部分構造を選択する方式を適用することができ

る。また、文書の最も外側の構造を始点として編集箇所記述要素とマッチする全ての部分構造を求める方式は、例えば、マッチする全ての組合せを求めたいが、処理の高速化も狙いたい場合などに適用することができる。例えば前記発表論文情報から一覧表Aの作成に必要な部分構造を抽出する例では、編集箇所記述として、

＜論文情報＞（％＜題名＞，％＜氏名＞）

といったように、編集箇所記述を、入力文書の最も外側のタグから順に記述することにより、編集箇所記述の先頭要素を、入力文書中の全ての部分構造を対象としてマッチを試みる処理などを省くことができ、高速化が図れる。

#### 【0086】

次に、実施例2を、図26のように、発表論文情報2から投稿先学会名と論文の題名を抽出して、一覧表Bを作成する例を用いて説明し、発明が解決しようとする課題の（2－3）が解決することを示す。

#### 【0087】

図26において、入力文書2601、入力文書2602は発表論文情報2であり、出力文書2604が作成したい一覧表Bである。また、図26の入力文書2601をSGML形式で記述した例が図27の構造化文書2701であり、図26の入力文書2602をSGML形式で記述した例が図27の構造化文書2702であり、図26の出力文書2604をSGML形式で記述した例が図27の出力文書2703である。

#### 【0088】

入力文書として図27の構造化文書2701および構造化文書2702の2つが与えられている例を説明する。

#### 【0089】

また、発表論文情報2から一覧表Bの作成に必要な部分構造を抽出するための編集箇所記述として、

％＜学会名＞＆％＜題名＞

が文書処理記述手段に記述されているとする。なお、以降では、前記編集箇所記述を編集箇所記述Bと呼ぶことにする。

#### 【0090】

また、文書出力手段による加工方法は、次の通りであるとする。なお、以降では、下記文書出力手段の加工方法を、加工方法Bと呼ぶことにする。

#### 【0091】

1. 編集箇所記述Bの1つめの引用指定子と

対応する部分構造

2. 編集箇所記述Bの2つめの引用指定子と

対応する部分構造

を文書蓄積手段からそれぞれ1つずつ抽出して

前記順番に並べたものを<表要素>というタグで

囲んで表要素を作り、文書蓄積手段中の全て

の部分構造に対して前記処理を実行して得られた

表要素を並べて、

全体を<一覧表B>というタグで囲んで出力する。

#### 【0092】

以上の条件の下での動作を説明する。

#### 【0093】

構造化文書入力手段101が、1つめの入力文書として、構造化文書2701を読み込む(ステップ2201)。文書編集処理実行手段104は、文書処理記述手段102から前記編集箇所記述Bを取り出して(ステップ2202)、構造化文書2701に対する抽出処理を実行する(ステップ2203)。

#### 【0094】

抽出処理では、OR接続子を区切りとして、編集箇所記述Bを複数の記述に分解する(ステップ2301)。編集箇所記述B中にはOR接続子はないため、分解された記述は、編集箇所記述B全体からなるもの1つだけとなる。前記編集箇所記述Bを処理対象として(ステップ2302)、AND接続子の処理を実行する(ステップ2303)。

#### 【0095】

AND接続子の処理では、AND接続子を区切りとして編集箇所記述Bを複数の記述に分解する(ステップ2401)。まず、区切った記述の1つ%<学会名>を処

理対象として選び(ステップ2402)、マッチ処理を実行する(ステップ2403)。

#### 【0096】

マッチ処理は、前述の第1の実施の形態の実施例1で詳細に説明したため、要点のみ示す。マッチ処理では、編集箇所記述の先頭要素である%<学会名>と構造化文書2701の全ての部分構造とのマッチを試み、結果として、マッチは成立し、編集箇所記述Bの1つめの引用指定子に対応する部分構造として、

<学会名>

知能研究会

</学会名>

が抽出され、文書蓄積手段103に蓄積される。

#### 【0097】

構造化文書2701と%<学会名>とのマッチ処理が完了すると、AND接続子の処理に戻る。AND接続子の処理では、処理すべき記述%<題名>が残っているため(ステップ2404)、次に、%<題名>を処理対象として選び(ステップ2402)、マッチ処理を実行する(ステップ2403)。

#### 【0098】

マッチ処理は、前述の第1の実施の形態の実施例1で詳細に説明したため、要点のみ示す。%<題名>を処理対象としたマッチ処理では、前述の%<学会名>を処理対象としたマッチ処理を実行した時とは別に、新たに、編集箇所記述の先頭要素である%<題名>と構造化文書2701の全ての部分構造とのマッチを試みる。その結果、マッチは成立し、編集箇所記述Bの2つめの引用指定子に対応する部分構造として、

<題名>

××に関する研究

</題名>

が抽出され、文書蓄積手段103に蓄積される。

#### 【0099】

構造化文書2701と%<題名>とのマッチ処理が完了すると、AND接続子の処理に戻る。AND接続子の処理では、分解した記述に対する処理は全て完了した

ため(ステップ2404)、前記編集箇所記述Bと入力文書とのマッチの状況を調べる(ステップ2405)。前記編集箇所記述BのAND接続子で区切られた各記述は全てマッチが成立したため、文書蓄積手段の削除は行わず、AND接続子の処理を終了する。

#### 【0100】

AND接続子の処理が完了すると、抽出処理に戻る。抽出処理では、分解した記述に対する処理は全て完了したため(ステップ2304)、前記編集箇所記述Bと入力文書とのマッチの状況を調べる(ステップ2305)。前記編集箇所記述Bはマッチが成立したため、文書蓄積手段の削除は行わず、抽出処理を終了する。

#### 【0101】

抽出処理が完了すると、全体の処理に戻る。全体の処理では、構造化文書2701に対して処理すべき編集箇所記述はもうないため(ステップ2204)、構造化文書入力手段101が、2つめの入力文書として、構造化文書2702を読み込む(ステップ2201)。文書編集処理実行手段104は、文書処理記述手段102から前記編集箇所記述Bを取り出して(ステップ2202)、構造化文書2702に対する抽出処理を実行する(ステップ2203)。

#### 【0102】

構造化文書2702に対する抽出処理は、前記構造化文書2701に対する抽出処理とほぼ同じであるため、詳細な説明は省き、結果だけを示す。処理の結果、前記編集箇所記述Bの1つめの引用指定子に対応する部分構造として、

<学会名>

情報研究会

</学会名>

が抽出され、また、前記編集箇所記述Aの2つめの引用指定子に対応する部分構造として、

<題名>

〇〇に関する研究

</題名>

が抽出され、文書蓄積手段103に蓄積され、抽出処理を終了する。

### 【0103】

抽出処理が完了すると、全体の処理に戻る。全体の処理では、構造化文書902に対して処理すべき編集箇所記述はもうなく(ステップ2204)、また、全ての構造化文書に対しての抽出処理が完了したため(ステップ2205)、文書出力手段105が、文書蓄積手段103に蓄積された部分構造を加工する(ステップ2206)。

### 【0104】

この時点で文書蓄積手段103に蓄積されている部分構造を、文書蓄積手段103の格納形式の一例を用いて以下に示す。

### 【0105】

編集箇所記述Aの1つめの引用指定子に対応する

部分構造

<学会名>

知能研究会

</学会名>

<学会名>

情報研究会

</学会名>

編集箇所記述Aの2つめの引用指定子に対応する

部分構造

<題名>

××に関する研究

</題名>

<題名>

〇〇に関する研究

</題名>

文書出力手段105は、これらの部分構造から、前記加工方法Bにより、図27の出力文書2703のような一覧表Bを作成し、出力する(ステップ2206)。

#### 【0106】

以上のように、本発明では、順不同の並列関係を表すAND接続子を記述することができ、また、例えば前記例でのマッチ処理における、編集箇所記述の先頭要素の処理のように、AND接続子で区切られた各記述は、それぞれ、構造化文書の全ての部分構造を開始点としてマッチが試みられ、AND接続子の処理のステップ2405、ステップ2406のように、区切った記述の全てのマッチが成功した組合せのみ、文書蓄積手段103に保持されるため、AND接続子の前後の記述が順不同でマッチして部分構造の抽出を実行することができる。

#### 【0107】

次に、実施例3として、括弧、および、OR接続子の解釈動作について示す。

#### 【0108】

例えば図28のようにSGML記述された論文、すなわち、

1. <図>および</図>というタグで囲まれた

図を表す部分構造と文書データとが、

<第一章>および</第一章>というタグで

囲まれた第一章を表す部分構造

2. <図>および</図>というタグで囲まれた

図を表す部分構造と文書データとが、

<第二章>および</第二章>というタグで

囲まれた第二章を表す部分構造

をこの順序で並べて全体を<論文>および</論文>というタグで囲んで表現したものから、第一章の図および第二章の図を抽出する動作を説明する。

#### 【0109】

入力文書として、図28の構造化文書2801が与えられている例を説明する。また、論文から第一章の図および第二章の図を抽出するための編集箇所記述として、

(<第一章> | <第二章>) %<図>

が文書処理記述手段に記述されているとする。なお、以降では、前記編集箇所記述を編集箇所記述Cと呼ぶことにする。

#### 【0110】

また、文書出力手段は、文書蓄積手段103の内容をそのまま出力するとする。

#### 【0111】

以上の条件の下での動作を簡単に説明する。

#### 【0112】

構造化文書入力手段101が構造化文書2801を読み込み(ステップ2201)、文書編集処理実行手段104が文書処理記述手段102から前記編集箇所記述Cを取り出して(ステップ2202)、構造化文書2801に対する抽出処理を実行する(ステップ2203)。

#### 【0113】

抽出処理では、編集箇所記述CにはOR接続子があるが、括弧で囲まれているため、編集箇所記述Cを複数の記述に分解することができず、前記編集箇所記述Cを処理対象として(ステップ2302)、AND接続子の処理を実行する(ステップ2303)。

#### 【0114】

AND接続子の処理では、編集箇所記述CにはAND接続子が含まれていないため、前記編集箇所記述Cを処理対象として(ステップ2402)、マッチ処理を実行する(ステップ2403)。

#### 【0115】

マッチ処理では、前記編集箇所記述Cの先頭要素(＜第一章＞|＜第二章＞)に着目し(2501)、(＜第一章＞|＜第二章＞)の前後の括弧をはずして(ステップ2503)、＜第一章＞|＜第二章＞に対する抽出処理を実行する(ステップ2504)。

#### 【0116】

抽出処理では、＜第一章＞|＜第二章＞をOR接続子で分解し(ステップ2301)、＜第一章＞を処理対象として(ステップ2302)、AND接続子の処理を実行する(ステップ2303)。

#### 【0117】

以降のAND接続子の処理の詳細は省略し、結果として、

a. ＜第一章＞と構造化文書2801とのマッチは成立

a-1. 抽出した部分構造はなし

a - 2. 構造化文書2801の着目している

部分構造は<第一章>

となり、AND接続子の処理を終了して、抽出処理に戻る。

#### 【0118】

抽出処理では、分解した2つめの記述<第二章>を処理対象として、AND接続子の処理を実行する。<第二章>を処理対象としたAND接続子の処理結果は、

b. <第二章>と構造化文書2801とのマッチは成立

b - 1. 抽出した部分構造はなし

b - 2. 構造化文書2801の着目している

部分構造は<第二章>

となり、抽出処理へ戻る。

#### 【0119】

抽出処理では、ORで区切った記述の少なくとも1つのマッチが成立しているため(ステップ2305)、<第一章> | <第二章>のマッチが成立したとして、抽出処理を終了し、マッチ処理へ戻る。

#### 【0120】

マッチ処理では、着目点の要素(<第一章> | <第二章>)と、次の要素%<図>との間の接続子が階層接続子であるため、前記AND接続子の処理結果a - 2. およびb - 2. より、前記a. のマッチに対しては、<第一章>の内側が構造化文書2801の現在着目している部分構造となり、前記b. のマッチに対しては、<第二章>の内側が構造化文書2801の現在着目している部分構造となる(ステップ2507)。

#### 【0121】

前記編集箇所記述Cの着目点が次の要素%<図>に移され(ステップ2508)、前記構造化文書2801中の現在着目している2箇所の部分構造とのマッチが試みられる(ステップ2505)。

#### 【0122】

その結果、前記a. のマッチに対しては、

<図>

図 1

</図>

が抽出され、また、前記 b. のマッチに対しては、

<図>

図 2

</図>

が抽出され、文書蓄積手段103に蓄積される(ステップ2511)。

### 【0 1 2 3】

最後に、文書出力手段105が、前記抽出された部分構造を出力し(ステップ2206)、全体の処理が終了する。

### 【0 1 2 4】

図 2 は本発明の第 2 の実施の形態の構成図である。図 2 を参照すると、本実施の形態の文書編集方式は、第 1 の実施の形態の構成に加えて、構造化文書や部分構造を蓄積する出力文書蓄積手段106と、動作記述を解釈し、文書の追加、文書の結合、文書数のカウントなどの編集処理を実行して、構造化文書や部分構造を前記出力文書蓄積手段に蓄積する動作記述解釈手段107と、読み込まれた各文書に対しての抽出処理を開始する前に実行する動作を記述した文書毎前処理記述手段108と、読み込まれた各文書に対しての前記抽出処理が完了する毎に実行する動作を記述した文書毎後処理記述手段109と、読み込まれた各文書に対しての前記抽出処理が行われる前に、前記文書毎前処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する文書毎前処理実行手段110と、読み込まれた各文書に対しての前記抽出処理が完了する毎に、前記文書毎後処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する文書毎後処理実行手段111と、読み込まれた文書の最初の文書に対する前記抽出処理を開始する前に実行する動

作を記述した初期処理記述手段112と、  
読み込まれた全ての文書に対する前記抽出処理が完了した後に実行する動作を記述した終了時処理記述手段113と、  
読み込まれた文書の最初の文書に対する前記抽出処理が行われる前に、前記初期処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する初期処理実行手段114と、  
読み込まれた全ての文書に対する前記抽出処理が完了した後に、前記終了処理記述手段に記述された動作記述を前記動作記述解釈手段へ通知する終了時処理実行手段115と、  
タグ名変更指示や構造削除指示や文字列変更指示や文字列削除指示を受け付けて各処理を実行する文書修正手段116と、  
構造化文書や部分構造を指定されたキーに従ってソートする文書ソート手段117とをさらに有する。

#### 【0125】

次に、動作について図23、図29、図30、図31、図32、図33、図34、図35を参照して説明する。

#### 【0126】

図29は、第2の実施の形態の全体処理動作を示す流れ図である。

#### 【0127】

初期処理実行手段114が初期処理を実行(ステップ2901)後、構造化文書入力手段101は文書編集の対象となる複数の構造化文書を1つ読み込む(ステップ2902)。さらに、文書毎前処理実行手段110が文書毎前処理を実行(ステップ2903)後、文書編集処理実行手段104は、文書処理記述手段102から編集処理記述を1つ取り出して(ステップ2904)、抽出・動作記述解釈処理を実行し(ステップ2905)、次に処理すべき編集処理記述がある場合(ステップ2906)には、編集処理記述を1つ取り出す処理2904に戻り、また、処理すべき編集処理記述がない場合には、文書毎後処理実行手段111が文書毎後処理を実行(ステップ2907)後、次に処理すべき構造化文書があるかを調べる(ステップ2908)。全ての構造化文書に対する処理がまだ完了していない場合には、ステップ2902の文書を1つ読み込む処理に戻り、ま

た、全ての構造化文書に対する処理が完了した場合には、終了時処理実行手段115が終了時処理を実行(ステップ2909)後、文書出力手段105が、出力文書蓄積手段106に蓄積された部分構造をそのまま、あるいは加工して出力する(ステップ2910)。

#### 【0128】

前記初期処理における動作フローを、図30を用いて説明する。

#### 【0129】

初期処理では、初期処理実行手段114が、初期処理記述手段112から初期処理記述を読み込み(ステップ3001)、前記読み込んだ初期処理記述を動作記述解釈手段107に通知して、動作記述解釈処理を実行する(ステップ3002)。

#### 【0130】

前記文書毎前処理における動作フローを、図31を用いて説明する。

#### 【0131】

文書毎前処理では、文書毎前処理実行手段110が、文書毎前処理記述手段108から文書毎前処理記述を読み込み(ステップ3101)、前記読み込んだ文書毎前処理記述を動作記述解釈手段107に通知して、動作記述解釈処理を実行する(ステップ3102)。

#### 【0132】

前記文書毎後処理における動作フローを、図32を用いて説明する。

#### 【0133】

文書毎後処理では、文書毎後処理実行手段111が、文書毎後処理記述手段109から文書毎後処理記述を読み込み(ステップ3201)、前記読み込んだ文書毎後処理記述を動作記述解釈手段107に通知して、動作記述解釈処理を実行する(ステップ3202)。

#### 【0134】

前記終了時処理における動作フローを、図33を用いて説明する。

#### 【0135】

終了時処理では、終了時処理実行手段115が、終了時処理記述手段113から終了時処理記述を読み込み(ステップ3301)、前記読み込んだ終了時処理記述を動作記

述解釈手段107に通知して、動作記述解釈処理を実行する(ステップ3302)。

#### 【0136】

前記抽出・動作記述解釈処理における動作フローを、図34を用いて説明する。

#### 【0137】

抽出・動作記述解釈処理では、文書編集処理実行手段104が、前記ステップ2904の編集処理記述を1つ取り出す処理で取り出された編集処理記述の編集箇所記述を用いて抽出処理を実行(ステップ3401)し、その結果、マッチが成立した場合(ステップ3402)には、前記抽出処理によって抽出された部分構造を文書蓄積手段103より取り出して、前記取り出した編集処理記述中の動作記述とともに動作記述解釈手段107に通知して、動作記述解釈処理を実行する(ステップ3403)。また、前記抽出処理を実行した結果、マッチが成立しなかった場合には、動作記述解釈処理を実行せず、抽出・動作記述解釈処理を終了する。

#### 【0138】

前記動作記述解釈処理における動作フローを、図35を用いて説明する。

#### 【0139】

動作記述解釈処理では、動作記述解釈手段107が、前記通知された動作記述を解釈して、タグ名変更指示、構造削除指示、文字列変更指示、あるいは、文字列削除指示などの文書修正指示であった場合(ステップ3501)には、前記文書修正指示と、指定された部分構造とを、文書修正手段116に通知し、修正結果を得る(ステップ3502)。また、前記通知された動作記述が文書ソート指示であった場合(ステップ3503)には、前記文書ソート指示と、指定された部分構造とを、文書ソート手段117に通知し、ソートした結果を得る(ステップ3504)。また、前記通知された動作記述が、文書修正指示でも文書ソート指示でもなかった場合には、前記動作記述に従い、一時変数への格納や演算、出力文書蓄積手段106への部分構造の格納などの動作実行処理を実行する(ステップ3505)。前記ステップ3502の文書修正処理、もしくは、前記ステップ3504の文書ソート処理、もしくは、前記ステップ3505の動作実行処理の実行後、まだ処理すべき動作記述がある場合(ステップ3506)には、動作記述解釈処理の先頭に戻る。また、処理すべき動作記述がもうない場合には、動作記述解釈処理を終了する。

#### 【0140】

なお、図34における抽出処理は、前記図23を用いて説明した抽出処理と同一である。

#### 【0141】

##### 【実施例】

次に、本発明の第2の実施の形態の実施例として、図2および図23、図29、図30、図31、図32、図33、図34、図35の構成による方式の動作を、具体例を用いて説明する。

#### 【0142】

例えば、図37のように、入力文書3701や入力文書3702の論文情報から文書編集方式3703により題名を抽出し、また、各題名の著者数を数えて、題名と著者人数とからなり、かつ、各著者人数を全て合計した合計人数が最後に添付された一覧表(以降、一覧表Cと呼ぶ)を出力文書3704として作成したいとする。

#### 【0143】

入力文書3701、3702は、図9の構造化文書901、902のようなSGML形式であるとする。

#### 【0144】

また、作成したい一覧表Cは、図38のようなSGML形式、すなわち、

1. <論文題名>および</論文題名>という

タグで囲まれた論文の題名を表す部分構造

2. <著者人数>および</著者人数>という

タグで囲まれた著者の人数を表す部分構造

をこの順序で並べて全体を<表要素>および</表要素>というタグで囲んだ部分構造を、さらに複数並べて、最後に、部分構造

3. <合計人数>および</合計人数>という

タグで囲まれた著者の合計人数を表す部分構造

が追加された部分構造の並びの全体を、<一覧表C>および</一覧表C>というタグで囲んだ形式で表現する。

#### 【0145】

また、文書処理記述手段には、次の編集処理記述D－1および編集処理記述D－2の2つの編集処理記述が格納されている。なお、変数daimei、および、変数countは、動作記述解釈手段107が内部で管理する変数である。変数daimeiは、部分構造を格納することのできる変数であり、ここでは、抽出した部分構造を格納するために用いている。また、変数countは、整数値を格納することのできる変数であり、ここでは、各論文毎の著者人数をカウントするために用いている。

編集処理記述D－1：次の編集箇所記述1と動作記述1の組から成る

編集箇所記述1：％＜題名＞

動作記述1：編集箇所記述1の1つめの引用  
指定子と対応する部分構造中の  
＜題名＞タグを＜論文題名＞タグに  
変更したものを、  
変数daimeiに格納する。

編集処理記述D－2：次の編集箇所記述2と動作記述2の組から成る

編集箇所記述2：％＜人数＞

動作記述2：編集箇所記述2の1つめの引用  
指定子と対応する部分構造の  
数を数え、  
変数countにその値を格納する。

#### 【0146】

また、初期処理記述手段には、次の初期処理記述1および初期処理記述2の2つのステップが格納されている。なお、変数total、および、変数tableは、動作記述解釈手段107が内部で管理する変数である。変数totalは、整数値を格納することのできる変数であり、ここでは、著者人数の合計をカウントするために用いている。また、変数tableは、部分構造を格納することのできる変数であり、ここでは、表要素の並びを作成するために用いている。

初期処理記述1：変数totalに0を代入する

初期処理記述2：変数tableの内容を空にする。

#### 【0147】

また、文書毎前処理記述手段には、次の文書毎前処理記述 1 が格納されている。

文書毎前処理記述 1 : 変数countに 0 を代入する。

#### 【 0 1 4 8 】

また、文書毎後処理記述手段には、次の文書毎後処理記述 1 および文書毎後処理記述 2 の 2 つのステップが格納されている。

文書毎後処理記述 1 : 変数totalの値を、変数countの値だけ増やす。

文書毎後処理記述 2 :

< 表要素 >

変数daimeiの内容

< 著者人数 >

変数countの内容

< / 著者人数 >

< / 表要素 >

の部分構造を構成し、

変数tableに追加する

また、終了時処理記述手段には、次の終了時処理記述 1 が格納されている。

終了時処理記述 1 :

< 一覧表 C >

変数tableの内容

< 合計人数 >

変数totalの内容

< / 合計人数 >

< / 一覧表 C >

の部分構造を構成し、

出力文書蓄積手段へ格納する。

#### 【 0 1 4 9 】

また、文書出力手段は、出力文書蓄積手段に格納された文書をそのまま出力する。

#### 【0150】

以上の条件の下での動作を説明する。

#### 【0151】

初期処理実行手段114が初期処理を実行し(ステップ2901)、初期処理記述手段112から、前記初期処理記述1および前記初期処理記述2を読み込み(ステップ3001)、動作記述解釈手段107へ通知して動作記述解釈処理を実行させる(ステップ3002)。

#### 【0152】

動作記述解釈手段107は、まず、初期処理記述1を解釈して動作実行処理を実行(ステップ3505)する。これにより、動作記述解釈手段107が内部で管理する変数totalの内容が0に初期化される。次に、動作記述解釈手段107は、初期処理記述2を解釈して動作実行処理を実行(ステップ3505)する。これにより、動作記述解釈手段107が内部で管理する変数tableの内容が空に初期化される。

#### 【0153】

初期処理の実行が完了すると、全体の処理に戻り、、構造化文書入力手段101が、1つめの入力文書として、構造化文書901を読み込む(ステップ2902)。

#### 【0154】

文書毎前処理実行手段110が文書毎前処理を実行し(ステップ2903)、文書毎前処理記述手段108から、前記文書毎前処理記述1を読み込み(ステップ3101)、動作記述解釈手段107へ通知して動作記述解釈処理を実行させる(ステップ3102)。

#### 【0155】

動作記述解釈手段107は、前記文書毎前処理記述1を解釈して動作実行処理を実行する(ステップ3505)。これにより、動作記述解釈手段107が内部で管理する変数countの内容が0に初期化される。

#### 【0156】

文書毎前処理の実行が完了すると、全体の処理に戻り、文書編集処理手段104が、文書処理記述手段102から、前記編集処理記述D-1を取り出し(ステップ2904)、抽出・動作記述解釈処理を実行する(ステップ2905)。

#### 【0157】

抽出・動作記述解釈処理では、文書編集処理実行手段104が、前記取り出した編集処理記述D－1の編集箇所記述1を用いて、構造化文書901の抽出処理を実行する(ステップ3401)。抽出処理の実行の結果、マッチが成立し、編集箇所記述1の1つめの引用指定子に対応する部分構造として、

<題名>

××に関する研究

</題名>

が抽出される。マッチが成立したため(ステップ3402)、文書編集処理実行手段104は、前記取り出した編集処理記述D－1の動作記述1と前記抽出した部分構造とを動作記述解釈手段107に通知し、動作記述解釈処理を実行させる(ステップ3403)。

#### 【0158】

動作記述解釈処理では、動作記述解釈手段107が、与えられた動作記述1を解釈する。動作記述1には文書修正指示が含まれているため(ステップ3501)、動作記述解釈手段107は、文書修正手段116に、前記抽出した文書と、「<題名>を<論文題名>に変える」という修正指示とを通知し、文書修正処理を実行させる。文書修正処理の実行結果として、動作記述解釈手段107は、部分文書

<論文題名>

××に関する研究

</論文題名>

を受けとる。さらに、動作記述解釈手段107は、動作記述1の残りの処理を解釈して、動作実行処理を行い(ステップ3505)、前記受けとった部分文書を、動作記述解釈手段107が内部で管理する変数daimeiに格納する。

#### 【0159】

動作記述解釈処理が完了し、さらに、抽出・動作記述解釈処理が完了して、全体の処理に戻り、次の編集処理記述D－2を用いた抽出・動作記述解釈処理が同様に実行される。

#### 【0160】

編集処理記述D－2を用いた抽出・動作記述解釈処理では、編集箇所記述2と

のマッチにより、

<氏名>

○× 太郎

</氏名>

<氏名>

×○ 花子

</氏名>

の2つの部分構造が抽出され、動作記述2の解釈によって、動作記述解釈手段107が内部で管理する変数countに2が代入される。

#### 【0161】

文書処理記述手段102の全ての編集処理記述の処理が完了したため(ステップ2906)、文書毎後処理実行手段111が文書毎後処理を実行し(ステップ2907)、動作記述解釈手段107が内部で管理する変数totalの値が2となり、また、動作記述解釈手段107が内部で管理する変数tableの内容が、

<表要素>

<論文題名>

××に関する研究

</論文題名>

<著者人数>

2

</著者人数>

</表要素>

となる。

#### 【0162】

同様に、構造化文書902の処理が行われ、その結果、動作記述解釈手段107が内部で管理する変数totalの値に1が加えられて3となり、また、動作記述解釈手段107が内部で管理する変数tableに新たな部分構造が追加され、変数tableの内容は、

<表要素>

<論文題名>

××に関する研究

</論文題名>

<著者人数>

2

</著者人数>

</表要素>

<表要素>

<論文題名>

〇〇に関する研究

</論文題名>

<著者人数>

1

</著者人数>

</表要素>

となる。

【0163】

全ての入力文書に対する処理が完了したため(ステップ2908)、終了時処理実行手段115が終了時処理を実行し(ステップ2909)、その結果、

<一覧表C>

<表要素>

<論文題名>

××に関する研究

</論文題名>

<著者人数>

2

</著者人数>

<／表要素>

<表要素>

<論文題名>

〇〇に関する研究

<／論文題名>

<著者人数>

1

<／著者人数>

<／表要素>

<合計人数>

3

<／合計人数>

<一覧表C>

という一覧表Cが完成し、出力文書蓄積手段106に格納される。

【0164】

最後に、文書出力手段105が、前記出力文書蓄積手段106に格納された一覧表Cを出力する(ステップ2910)。

【0165】

なお、前記例では、動作記述解釈手段107は、動作記述1の文書修正指示を解釈して、文書修正手段116に文書修正処理を実行させたが、文書修正手段116を用意せずに、動作記述解釈手段107がステップ3505の動作実行処理の中で文書修正処理を行ってもよい。

【0166】

また、前記例にはなかったが、同様に、動作記述解釈手段107に通知された動作記述に文書ソート指示が含まれていた場合には、文書ソート手段117を用意して、動作記述解釈手段107が文書ソート手段117に、文書ソート処理を実行させる他に、文書ソート手段117を用意せず、動作記述解釈手段107が、ステップ3505の

動作実行処理の中で文書ソート処理を行ってもよい。

#### 【0167】

図3は本発明の第3の実施の形態の構成図である。図3を参照すると、本実施の形態の文書編集方式は、第2の実施の形態の構成に加えて、前記読み込まれた文書に対する削除要求を受け付け、前記削除要求を前記文書編集処理実行手段および前記動作記述解釈手段に通知する入力文書削除要求受信手段118をさらに有する。

#### 【0168】

次に、動作について図23、図30、図31、図32、図33、図34、図36を参照して説明する。

#### 【0169】

図36は、第3の実施の形態の全体処理動作を示す流れ図である。

#### 【0170】

初期処理実行手段114が初期処理を実行(ステップ3601)後、構造化文書入力手段101は文書編集の対象となる構造化文書の到着を待つ(ステップ3602)。また、入力文書削除要求受信手段118は入力文書削除要求を待つ(ステップ3603)。

#### 【0171】

構造化文書が到着した場合には、構造化文書入力手段101が到着した構造化文書を1つ読み込む(ステップ3604)。さらに、文書毎前処理実行手段110が文書毎前処理を実行(ステップ3605)後、文書編集処理実行手段104は、文書処理記述手段102から編集処理記述を1つ取り出して(ステップ3606)、抽出・動作記述解釈処理を実行し(ステップ3607)、次に処理すべき編集処理記述がある場合(ステップ3608)には、ステップ3606の編集処理記述を1つ取り出す処理に戻り、また、処理すべき編集処理記述がない場合には、文書毎後処理実行手段110が文書毎後処理を実行(ステップ3609)後、まだ構造化文書の到着を待つべきかを判定する(ステップ3610)。まだ構造化文書の到着を待つべきと判断した場合には、構造化文書の到着を待つ処理に戻り、また、全ての構造化文書に対する処理が完了したと判断した場合には、終了時処理実行手段115が終了時処理を実行(ステップ3611)後、文書出力手段105が、出力文書蓄積手段106に蓄積された部分構造をそのま

ま、あるいは加工して出力する(ステップ3612)。

#### 【0172】

また、入力文書削除要求受信手段118が入力文書削除要求を受信した場合には、文書編集処理実行手段104および動作記述解釈手段107へ削除要求を通知する。入力文書削除要求受信手段118から削除要求を通知された文書編集処理実行手段104は、削除要求のあった文書に関する部分構造を文書蓄積手段103から削除し(ステップ3613)、また、削除した部分構造を動作記述解釈手段107へ通知する。入力文書削除要求受信手段118から削除要求を通知され、また、文書編集処理実行手段104から削除した部分構造を通知された動作記述解釈手段107は、削除要求された文書名や削除した部分構造などの情報をもとに、動作記述解釈手段107の内部情報や出力文書蓄積手段106の内容を変更して、削除要求された文書が到着していない場合と等しい状態にする(ステップ3614)。

#### 【0173】

なお、図36における初期処理は、前記図30を用いて説明した初期処理と同一である。

#### 【0174】

また、図36における文書毎前処理は、前記図31を用いて説明した文書毎前処理と同一である。

#### 【0175】

また、図36における抽出・動作記述解釈処理は、前記図34を用いて説明した抽出・動作記述解釈処理と同一である。

#### 【0176】

また、図36における文書毎後処理は、前記図32を用いて説明した文書毎後処理と同一である。

#### 【0177】

また、図36における終了時処理は、前記図33を用いて説明した終了時処理と同一である。

#### 【0178】

#### 【実施例】

次に、本発明の第3の実施の形態の実施例として、図3および図23、図30、図31、図32、図33、図34、図36の構成による方式の動作を、第2の実施の形態の実施例に用いた例と同様の例で説明する。

【0179】

ただし、

1. 構造化文書901が到着
2. 構造化文書902が到着
3. 構造化文書901の削除要求が到着
4. さらに別の構造化文書が到着

の順番で、入力文書や要求が到着するものとする。

【0180】

まず、構造化文書901および構造化文書902が順次到着すると、第2の実施の形態の説明の場合と同様に処理が行われ、その結果、動作記述解釈手段107が内部で管理する変数totalの値は3となり、また、動作記述解釈手段107が内部で管理する変数tableの内容は、

<表要素>

<論文題名>

××に関する研究

</論文題名>

<著者人数>

2

</著者人数>

</表要素>

<表要素>

<論文題名>

〇〇に関する研究

</論文題名>

<著者人数>

</著者人数>

</表要素>

となる。また、文書蓄積手段103には、部分構造が、抽出した構造化文書毎に格納されており、

構造化文書901から抽出した部分構造

編集箇所記述 1 の 1 つめの引用指定子と対応する

部分構造

<題名>

××に関する研究

</題名>

編集箇所記述 2 の 1 つめの引用指定子と対応する

部分構造

<氏名>

○× 太郎

</氏名>

<氏名>

×○ 花子

</氏名>

構造化文書902から抽出した部分構造

編集箇所記述1の 1 つめの引用指定子と対応する

部分構造

<題名>

○○に関する研究

</題名>

編集箇所記述 2 の 1 つめの引用指定子と対応する

部分構造

<氏名>

〇〇 次郎

</氏名>

となっている。

【0181】

次に、構造化文書901の削除要求が到着すると(ステップ3603)、入力文書削除要求受信手段118が文書編集処理実行手段104および動作記述解釈手段107に、前記削除要求を通知する。前記削除要求を受けとった文書編集処理実行手段104は、前記文書蓄積手段103から、入力文書901に関する部分構造を削除し、文書蓄積手段103の内容は、

構造化文書902から抽出した部分構造

編集箇所記述 1 の 1 つめの引用指定子と対応する

部分構造

<題名>

〇〇に関する研究

</題名>

編集箇所記述 2 の 1 つめの引用指定子と対応する

部分構造

<氏名>

〇〇 次郎

</氏名>

となる。また、動作記述解釈手段107は、入力文書削除要求受信手段118から削除要求を受けとり、かつ、文書編集処理実行手段104から前記部分構造の削除を通知されると、削除された文書が到着していない場合と等しい状態とするため、動作記述解釈手段107が内部で管理する変数totalの値から2を減じて1とし、また、動作記述解釈手段107が内部で管理する変数tableの内容の一部を削除して、変数tableの内容を、

<表要素>

<論文題名>

〇〇に関する研究

＜／論文題名＞

＜著者人数＞

1

＜／著者人数＞

＜／表要素＞

とする。

#### 【0182】

以後、さらに別の入力文書が到着する度に、第2の実施の形態の説明の場合と同様の処理が行われ、全ての入力文書が到着すると、終了時処理が実行されて(ステップ3611)、出力文書蓄積手段106に一覧表Cが格納され、文書出力手段105が、前記一覧表Cを出力する(ステップ3612)。

#### 【0183】

なお、前記例では、構造化文書の到着を待つべきと判断した場合には、終了時処理の実行(ステップ3611)および文書出力処理の実行(ステップ3612)は行われないうが、1つの文書に対する処理が完了する度に、文書毎後処理(ステップ3610)の後で、毎回、終了時処理の実行(ステップ3611)および文書出力処理の実行(ステップ3612)を行ってもよい。この場合には、それまでに到着した文書に関して集計した結果が一覧表Cとして毎回出力される。例えば、前記の例で、

1. 構造化文書901が到着
2. 構造化文書902が到着
3. 構造化文書901の削除要求が到着

の順番に入力文書や要求が到着した場合には、図39に示すような一覧表Cが、出力文書3901、3902、3903の順に得られる。

#### 【0184】

また、前記削除要求を受けとった場合の動作記述解釈手段107の処理は、動作記述解釈手段107が内部で管理する変数などから、削除要求された文書に関する部分を削除する方法の他に、動作記述解釈手段107が内部で管理する変数などの内容を全て破棄した上で、削除要求のあった文書に関する部分構造の削除された文書蓄積手段103をもとに、今まで到着した文書が再び同一順序で到着したかの

ように全体処理を実行して、削除要求された文書が到着していない場合と等しい状態にしてもよい。

#### 【0185】

なお、以上の実施例で説明した例において、文書蓄積手段103は、例えば、図4に示すように、各引用指定子に対応する部分構造が、引用指定子毎に、マッチが成立した順に並べたデータをメモリ上もしくはディスク上などに蓄積したデータベースとして実現する。もしくは、文書蓄積手段103は、例えば、図5に示すように、抽出した構造化文書毎に部分構造を分け、さらに、引用指定子毎に分けて格納されていてもよい。

#### 【0186】

また、文書処理記述手段102は、例えば、図6に示すように、編集箇所記述1つ1つを別々の格納領域に格納し、メモリ上もしくはディスク上などに蓄積したデータベースとして実現する。もしくは、文書処理記述手段102は、例えば、図7に示すように、編集箇所記述と動作記述の組からなる編集処理記述1つ1つを別々の格納領域に格納して構成してもよい。

#### 【0187】

また、動作記述解釈手段107は、独自の動作記述を解釈して実行する構成の他に、一般に良く知られているプログラミング言語を動作記述とし、前記プログラミング言語のインタプリタを利用することも可能である。

#### 【0188】

また、編集箇所記述は、実施例で用いた記号によって記述したものでなくとも、記号と編集箇所記述の各構成要素との対応が明確であれば、本発明は適用可能である。さらに、実施例では、編集箇所記述を記号を用いて表現したが、編集箇所記述を図形によって表現することもでき、この場合も、本発明は適用可能である。

#### 【0189】

また、本実施の形態では、SGML形式の構造化文書を例として用いたが、本発明は、他の形式の構造化文書に対しても適用可能である。文書の構造を表現する手段は、SGML形式などのように、特殊な意味を持たせた文字列（タグ）で

なく、四角などの図形や色などでもよい。さらに、本発明は、入力文書が、特別な文字や記号で構造が表現されている文書でなくとも、例えば、記述内容などから構造化できる文書であれば、入力手段が文書の構造を解釈して、前記例と同様に編集処理することができる。出力文書も、特別な文字や記号で構造が表現されている文書に限られているわけではなく、例えば、「文書修正手段によって全てのタグを削除してから出力文書蓄積手段に出力文書を格納する」といった動作記述を与えることにより、構造を表現する特別な文字や記号のない文書を出力することができる。さらに、構造化文書が、参照ポインタによってネットワーク状に接続された複数の構造体の集合、あるいは、参照ポインタによってネットワーク状に接続された複数の文書の集合として構成されていても、本発明は適用可能である。

#### 【0190】

また、実施例では、読み込む文書も出力する文書も同一のSGML形式である例を用いたが、本発明は、入力文書と出力文書の記述形式が異なる場合でも適用可能である。

#### 【0191】

また、本発明では、入力文書の構造を表現する手段（タグなど）を識別するための最小限の規約（例えば、SGML形式の場合には、構造は開始タグと終了タグで囲むことによって表現し、開始タグは<で始まり>で終る文字列であり、終了タグは</で始まり>で終る文字列である、といった文書構造の表現形式や、省略可能なタグなど）が与えられていれば十分であり、本発明を構成するいずれの処理も、構造化文書を解釈するための構造定義（タグの意味のある構成を表す文法定義など）を必要としない。すなわち、（１）構造定義があいまいな構造化文書を入力あるいは出力とする場合や、（２）構造定義が繁雑に変更される構造化文書を入力あるいは出力とする場合、あるいは、前記「発明の属する技術分野」で説明した利用を目的とするシステムを設計あるいは作成あるいは運用する際に、（３）構造定義を設計する作業や、作成する作業や、保守する作業を省くと、システム構築や運用のための工数が著しく効率化される場合、などにおいて、本発明の文書編集方式は、構造定義の設計や作成や保守などの作業負担を減らす

ことができるという効果もある。

#### 【0192】

また逆に、構造化文書入力手段101や文書出力手段105で、入力文書や出力文書の文書構造のチェックを行ってもよい。

#### 【0193】

##### 【発明の効果】

以上述べたように、本発明には、例えば前記編集箇所記述Aのように、1つのパターンに複数の抽出箇所を指定することができ、しかも、例えば、部分構造の並びの順番という形で、各引用指定子に対応する部分構造の間のつながりを維持して部分構造を抽出することができ、さらに、構造を分解する処理を特別に用意することなく、例えば前記加工方法Aのように、抽出した部分構造を、引用指定子毎に別々に扱うことができるという効果がある。

#### 【0194】

また、例えば、第1の実施の形態の実施例2として示したように、本発明には、順不同の並列関係を表すAND接続子を記述し、また、処理することができるため、SGMLなどの形式で記述された構造化文書を扱う場合には、入力文書から抽出する部分構造を指定する編集箇所記述が簡潔に記述できるという効果がある。

#### 【0195】

また、本発明には、用途や要求に応じたマッチの実行方式を選択することにより、処理速度の高速化、使用する作業領域の大きさなどの要求にこたえることができるという効果もある。

##### 【図面の簡単な説明】

##### 【図1】

本発明の第1の実施の形態の構成図である。

##### 【図2】

本発明の第2の実施の形態の構成図である。

##### 【図3】

本発明の第3の実施の形態の構成図である。

**【図 4】**

本発明の文書蓄積手段の構成の 1 つを示す説明図である。

**【図 5】**

本発明の文書蓄積手段の構成の 1 つを示す説明図である。

**【図 6】**

本発明の文書処理記述手段の構成の 1 つを示す説明図である。

**【図 7】**

本発明の文書処理記述手段の構成の 1 つを示す説明図である。

**【図 8】**

本発明の文書編集方式の利用例の 1 つを示す説明図である。

**【図 9】**

本発明の文書編集方式の入力文書の一例を示す説明図である。

**【図 10】**

本発明の文書編集方式の出力文書の一例を示す説明図である。

**【図 11】**

従来の文書編集方式の動作の一例を示す説明図である。

**【図 12】**

従来の文書編集方式の動作の一例を示す説明図である。

**【図 13】**

本発明の文書タグのマッチの一例を示す説明図である。

**【図 14】**

本発明の文書文字パターンのマッチの一例を示す説明図である。

**【図 15】**

本発明のワイルドカードタグのマッチの一例を示す説明図である。

**【図 16】**

本発明の否定指定子の一例を示す説明図である。

**【図 17】**

本発明の引用指定子の一例を示す説明図である。

**【図 18】**

本発明の順序接続子の一例を示す説明図である。

【図 1 9】

本発明の階層接続子の一例を示す説明図である。

【図 2 0】

本発明のAND接続子の一例を示す説明図である。

【図 2 1】

本発明のOR接続子の一例を示す説明図である。

【図 2 2】

本発明の第1の実施の形態の全体処理動作を示す流れ図である。

【図 2 3】

本発明の方式の抽出処理の一例を示す流れ図である。

【図 2 4】

本発明の方式の抽出処理の一例を示す流れ図である。

【図 2 5】

本発明の方式のマッチ処理の一例を示す流れ図である。

【図 2 6】

本発明の文書編集方式の利用例の1つを示す説明図である。

【図 2 7】

本発明の文書編集方式の入力文書および出力文書の一例を示す説明図である。

【図 2 8】

本発明の文書編集方式の入力文書の一例を示す説明図である。

【図 2 9】

本発明の第2の実施の形態の全体処理動作を示す流れ図である。

【図 3 0】

本発明の方式の初期処理の一例を示す流れ図である。

【図 3 1】

本発明の方式の文書毎前処理の一例を示す流れ図である。

【図 3 2】

本発明の方式の文書毎後処理の一例を示す流れ図である。

**【図 3 3】**

本発明の方式の終了時処理の一例を示す流れ図である。

**【図 3 4】**

本発明の方式の抽出・動作記述解釈処理の一例を示す流れ図である。

**【図 3 5】**

本発明の方式の動作記述解釈処理の一例を示す流れ図である。

**【図 3 6】**

本発明の第 3 の実施の形態の全体処理動作を示す流れ図である。

**【図 3 7】**

本発明の文書編集方式の利用例の 1 つを示す説明図である。

**【図 3 8】**

本発明の文書編集方式の出力文書の一例を示す説明図である。

**【図 3 9】**

本発明の文書編集方式の出力文書の一例を示す説明図である。

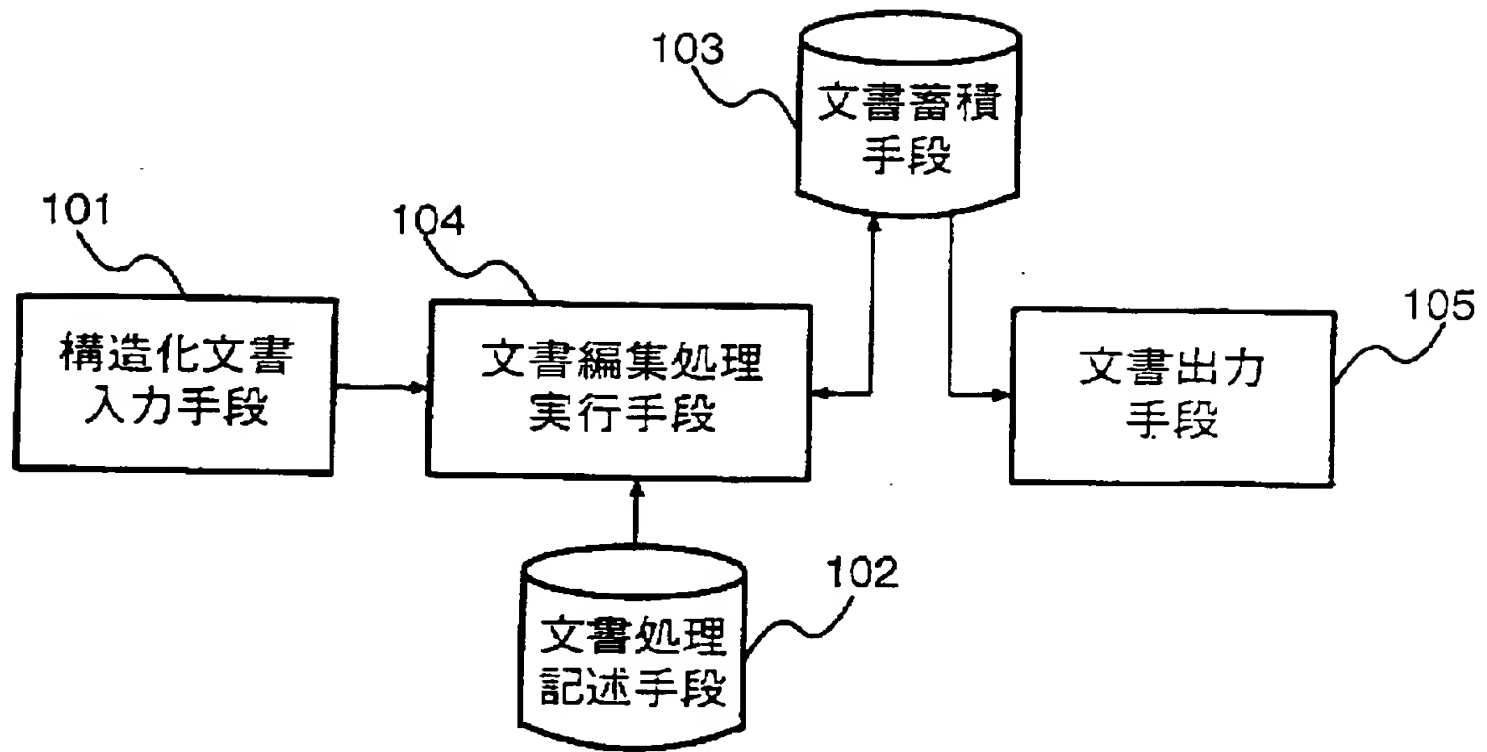
**【符号の説明】**

- 101 構造化文書入力手段
- 102 文書処理記述手段
- 103 文書蓄積手段
- 104 文書編集処理実行手段
- 105 文書出力手段
- 106 出力文書蓄積手段
- 107 動作記述解釈手段
- 108 文書毎前処理記述手段
- 109 文書毎後処理記述手段
- 110 文書毎前処理実行手段
- 111 文書毎後処理実行手段
- 112 初期処理記述手段
- 113 終了時処理記述手段
- 114 初期処理実行手段

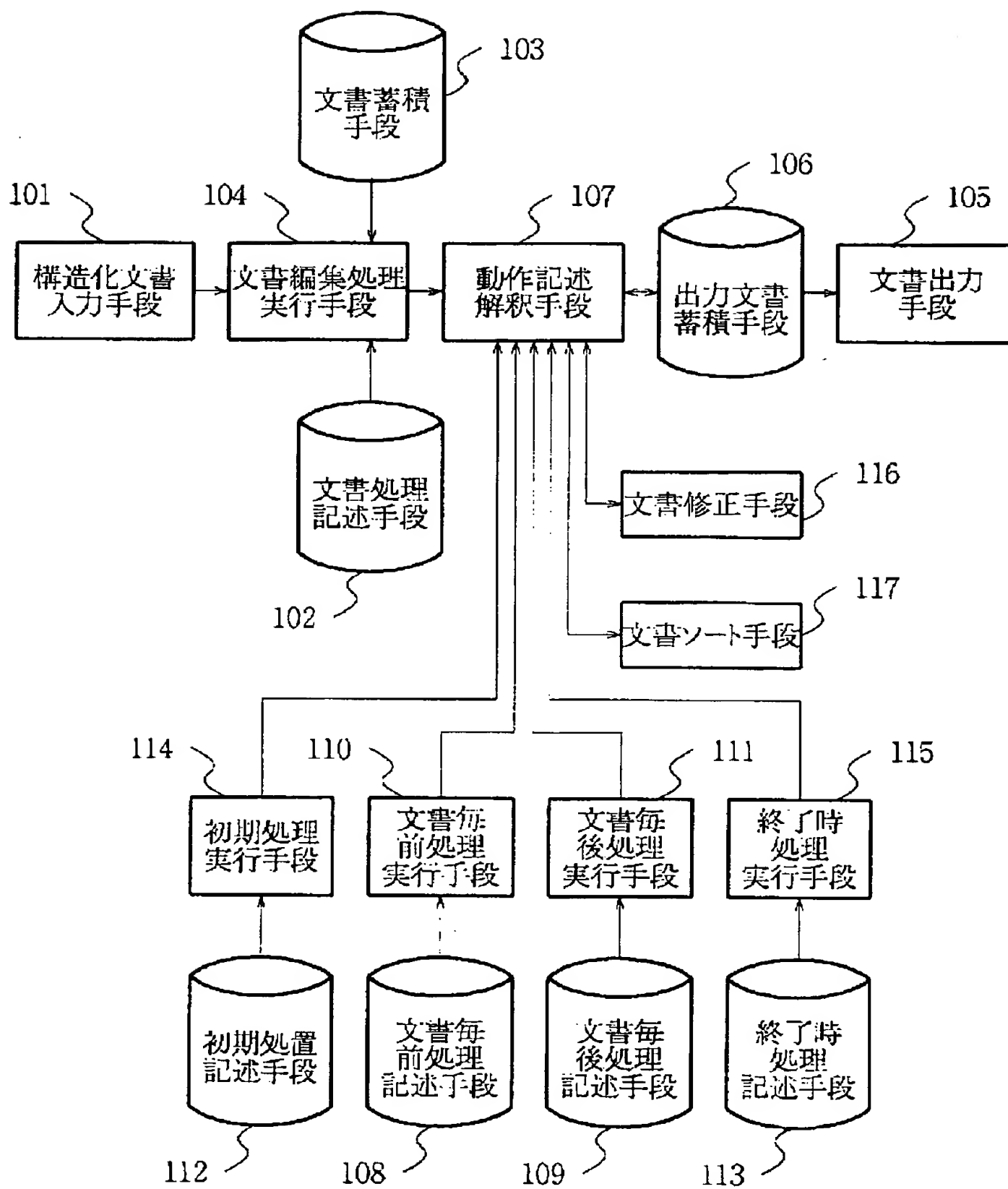
115 終了時処理実行手段  
 116 文書修正手段  
 117 文書ソート手段  
 118 入力文書削除要求受信手段  
 801、802、2601、2602、3701、3702 入力文書  
 901、902、1302、1402、1502、1602、1702、1802、1902、2002、2102、2701、  
 2702、2801 構造化文書  
 803、2603、3703 文書編集方式  
 804、1101、1201、2604、2703、3704、3901、3902、3903 出力文書  
 1102、1103 部分構造  
 1202 部分文書  
 1301、1401、1501、1601、1701、1801、1901、2001、2101 編集箇所記述

【書類名】      図面

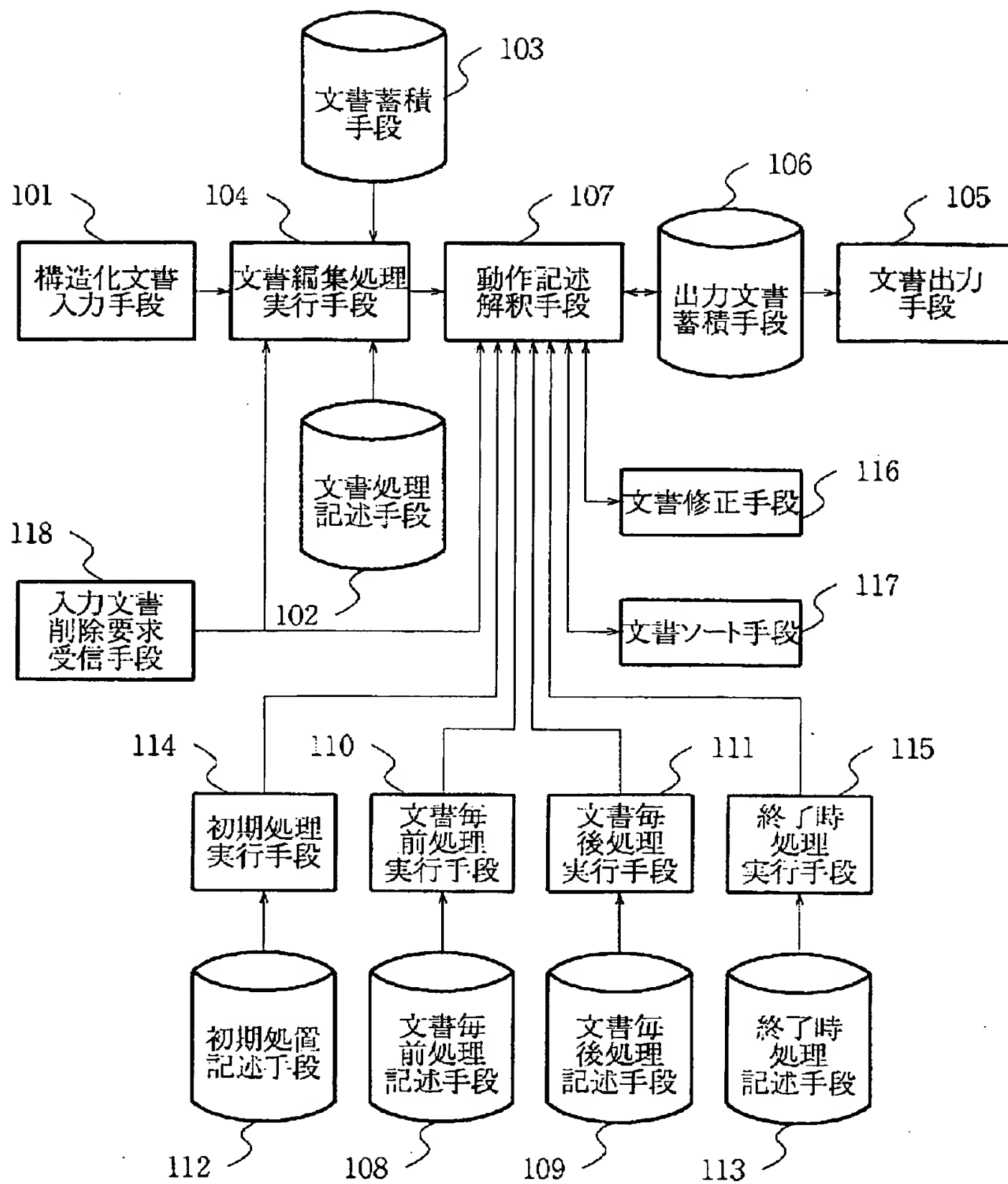
【図 1】

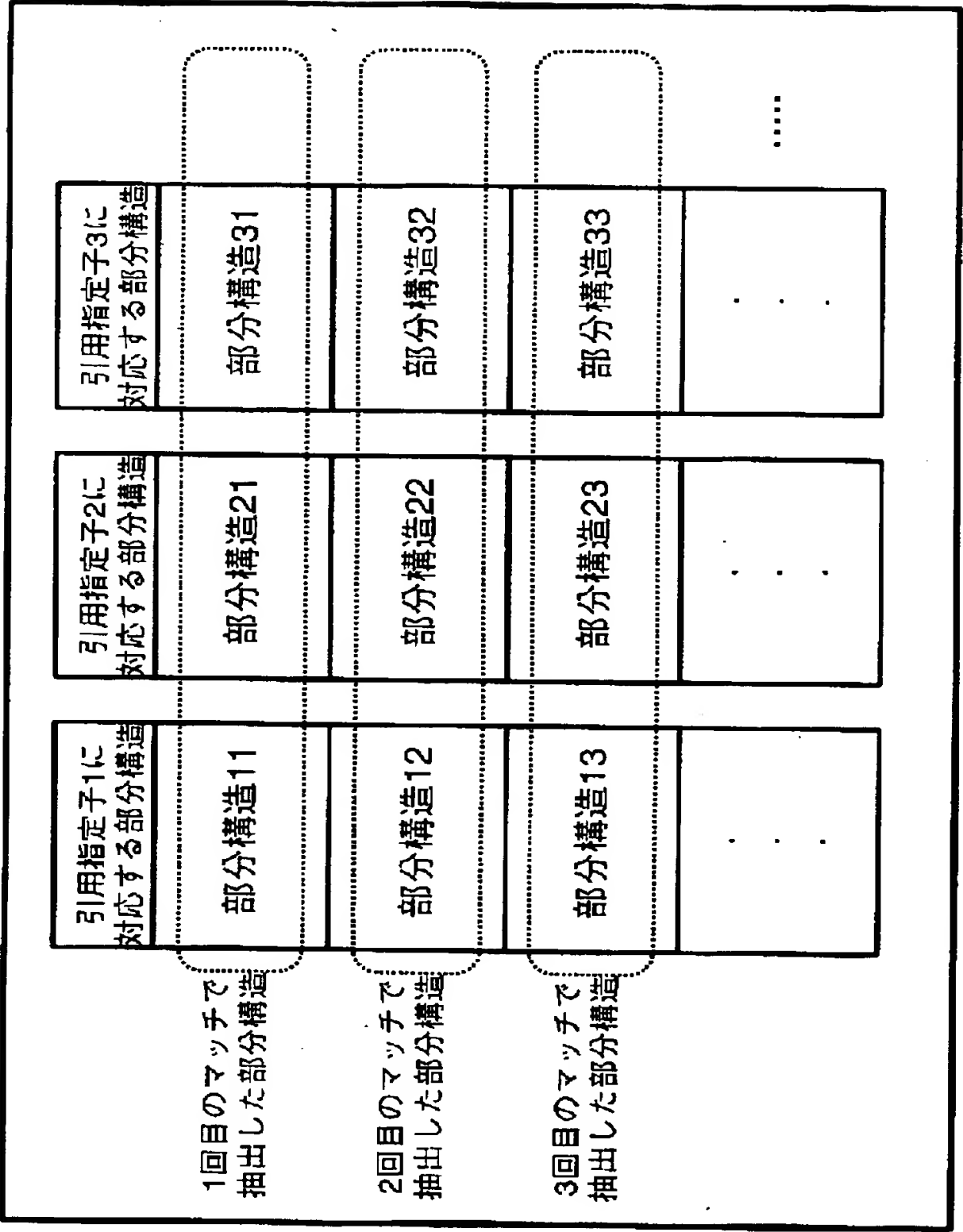


【図 2】



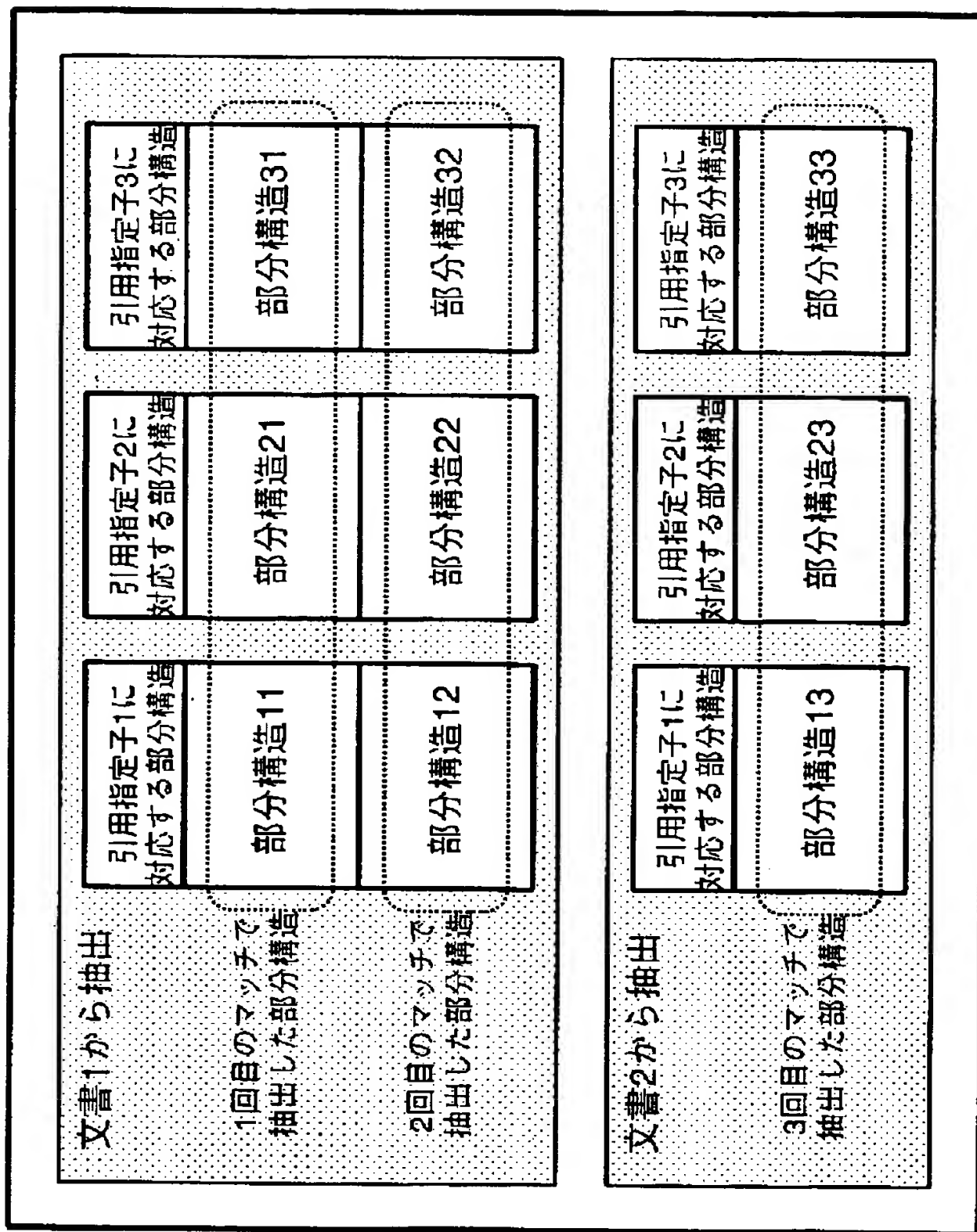
【図 3】





【図 4】

【図 5】

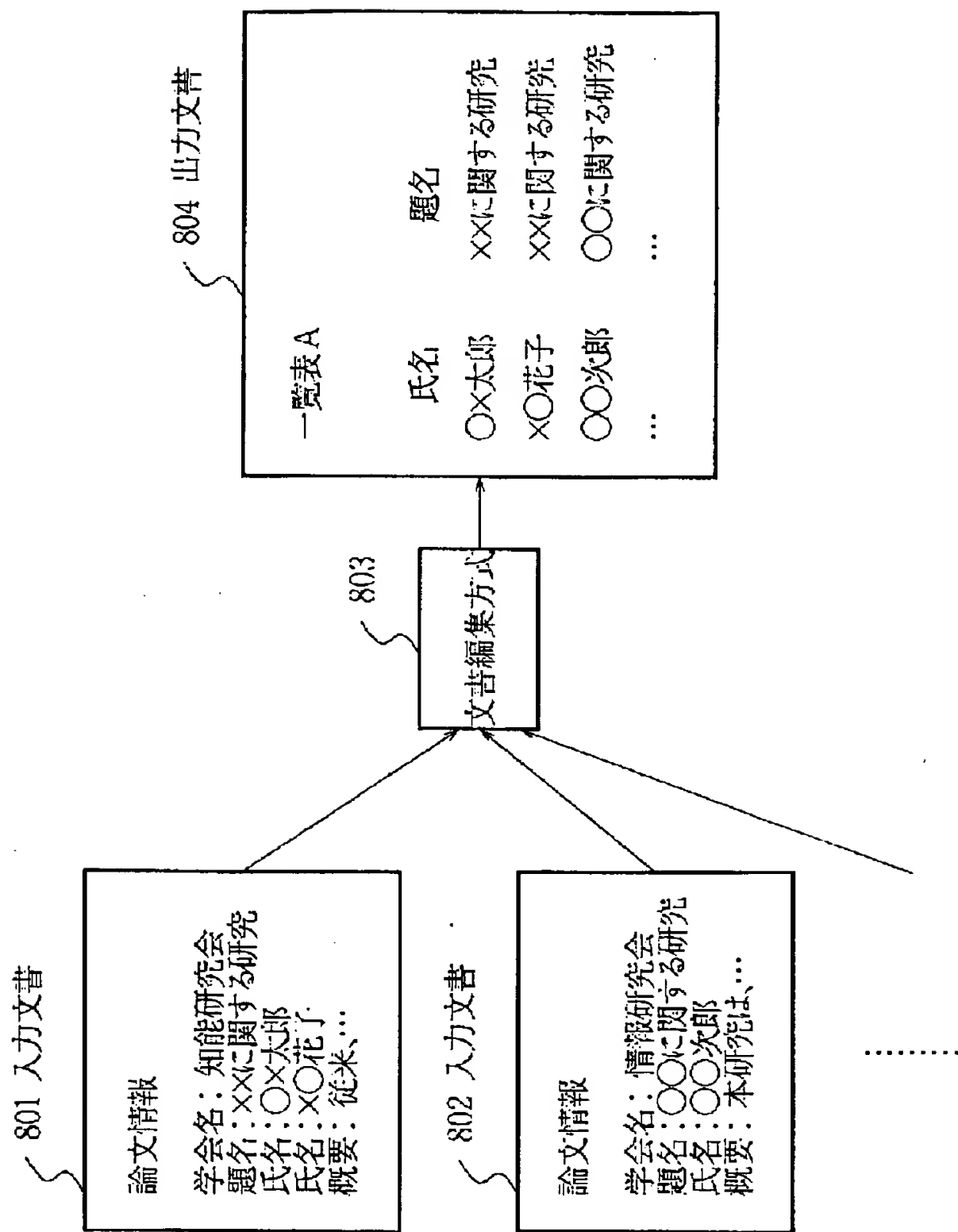


【図 6】

%<題名>,%<氏名>
%<学会名>&%<題名>
(<第一章> <第二章>)%<図>
.....
.
.
.
.
.

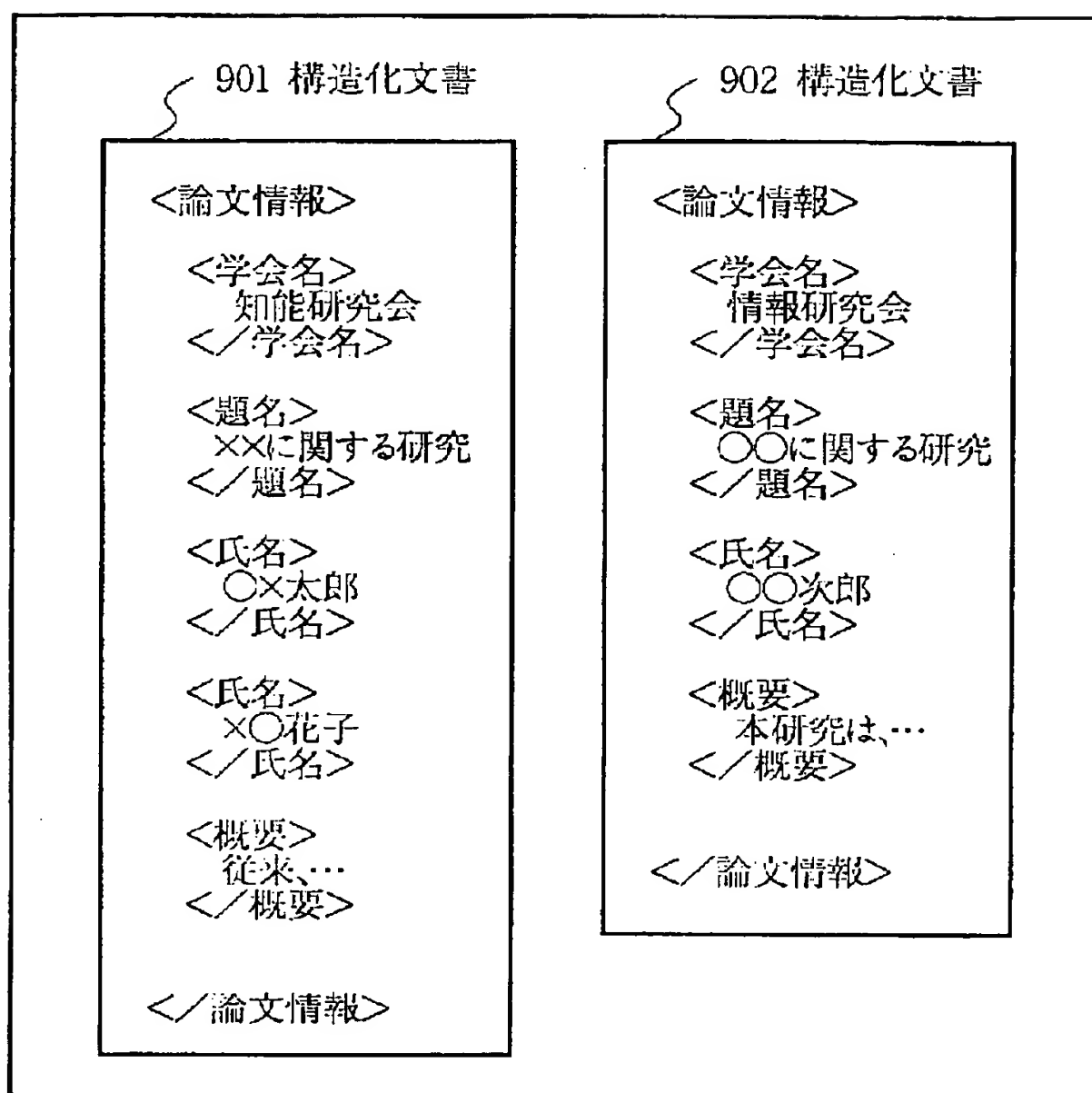
【図 7】

編集箇所記述	動作記述
%<題名>, %<氏名>	<p>(1)編集箇所記述の2つめの引用指定子と対応する部分構造</p> <p>(2)編集箇所記述の1つめの引用指定子と対応する部分構造</p> <p>を文書蓄積手段からそれぞれ1つずつ抽出して前記順番に並べたものを&lt;表要素&gt;というタグで囲んで表要素を作り、</p> <p>文書蓄積手段中の全ての部分構造に対して前記処理を実行して得られた表要素を並べて、全体を&lt;一覧表 A&gt;というタグで囲んで出力する。</p>
編集箇所記述	動作記述
%<学会名>&%<題名>	<p>(1)編集箇所記述の1つめの引用指定子と対応する部分構造</p> <p>(2)編集箇所記述の2つめの引用指定子と対応する部分構造</p> <p>を文書蓄積手段からそれぞれ1つずつ抽出して前記順番に並べたものを&lt;表要素&gt;というタグで囲んで表要素を作り、</p> <p>文書蓄積手段中の全ての部分構造に対して前記処理を実行して得られた表要素を並べて、全体を&lt;一覧表 B&gt;というタグで囲んで出力する。</p>



【図 8】

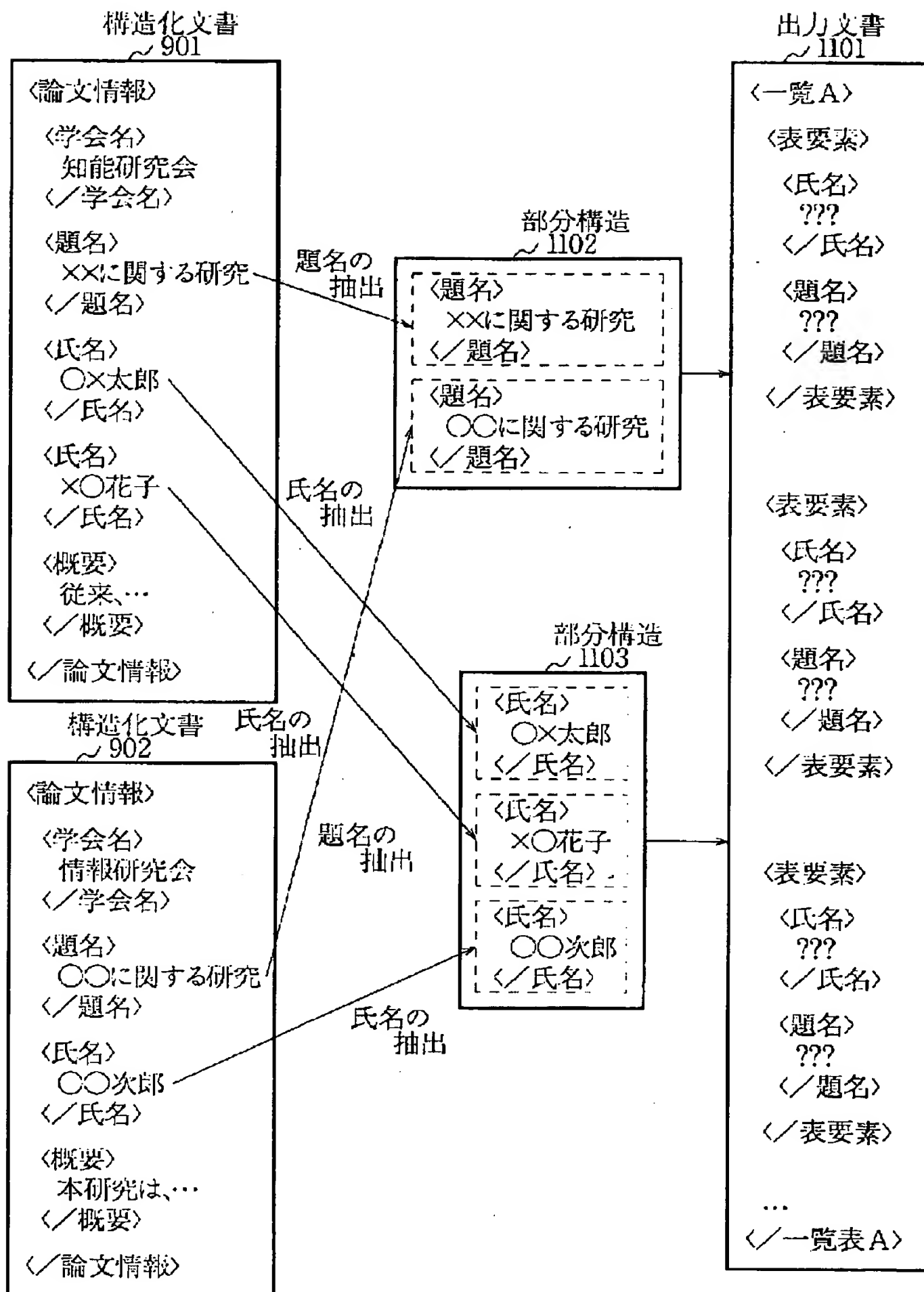
【図 9】



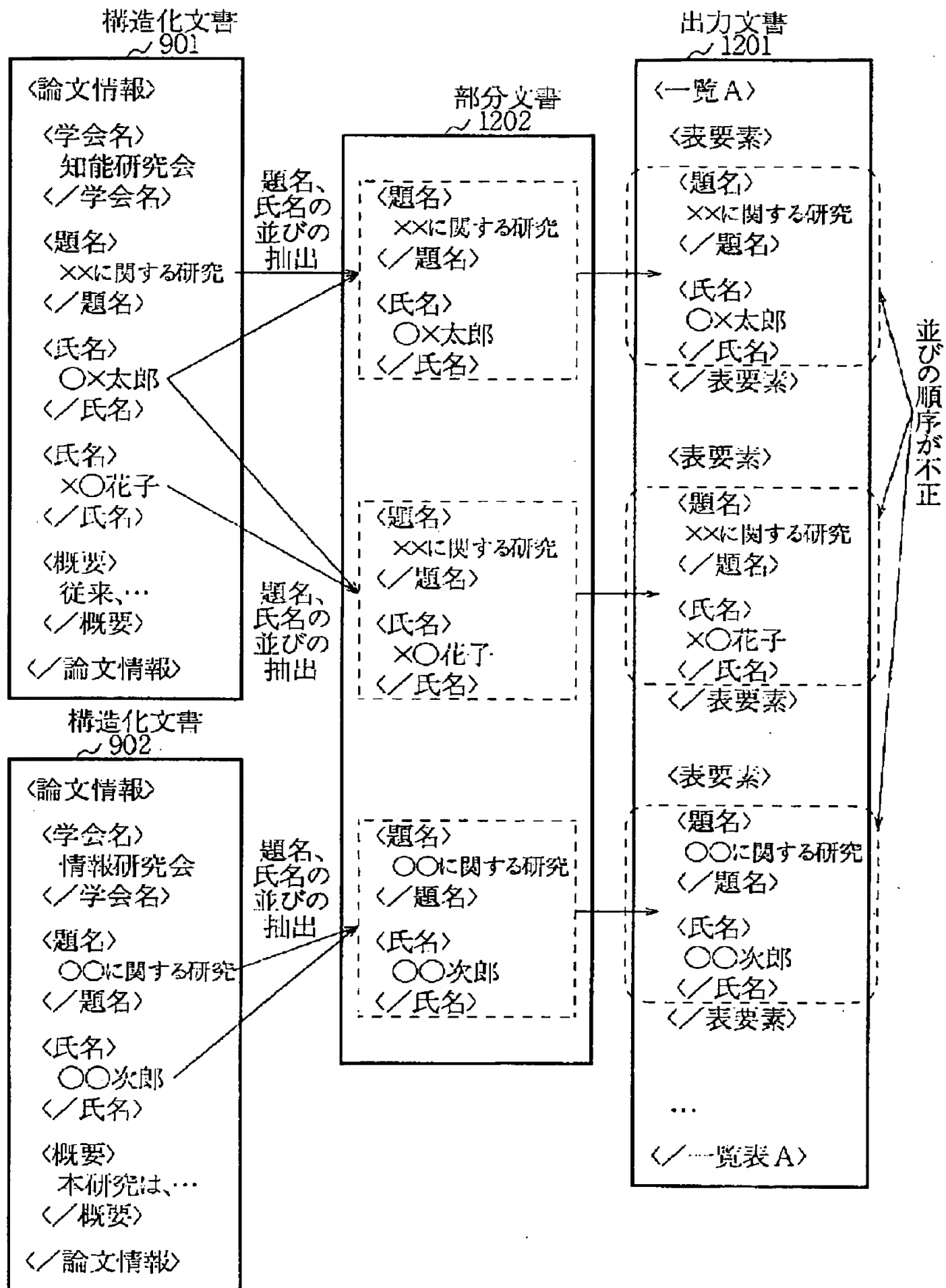
【図 1 0】

<一覧表A>	
<表要素>	
<氏名>	○× 太郎
</氏名>	
<題名>	××に関する研究
</題名>	
</表要素>	
<表要素>	
<氏名>	×○ 花子
</氏名>	
<題名>	××に関する研究
</題名>	
</表要素>	
<表要素>	
<氏名>	○○ 次郎
</氏名>	
<題名>	○○に関する研究
</題名>	
</表要素>	
.....	
</一覧表A>	

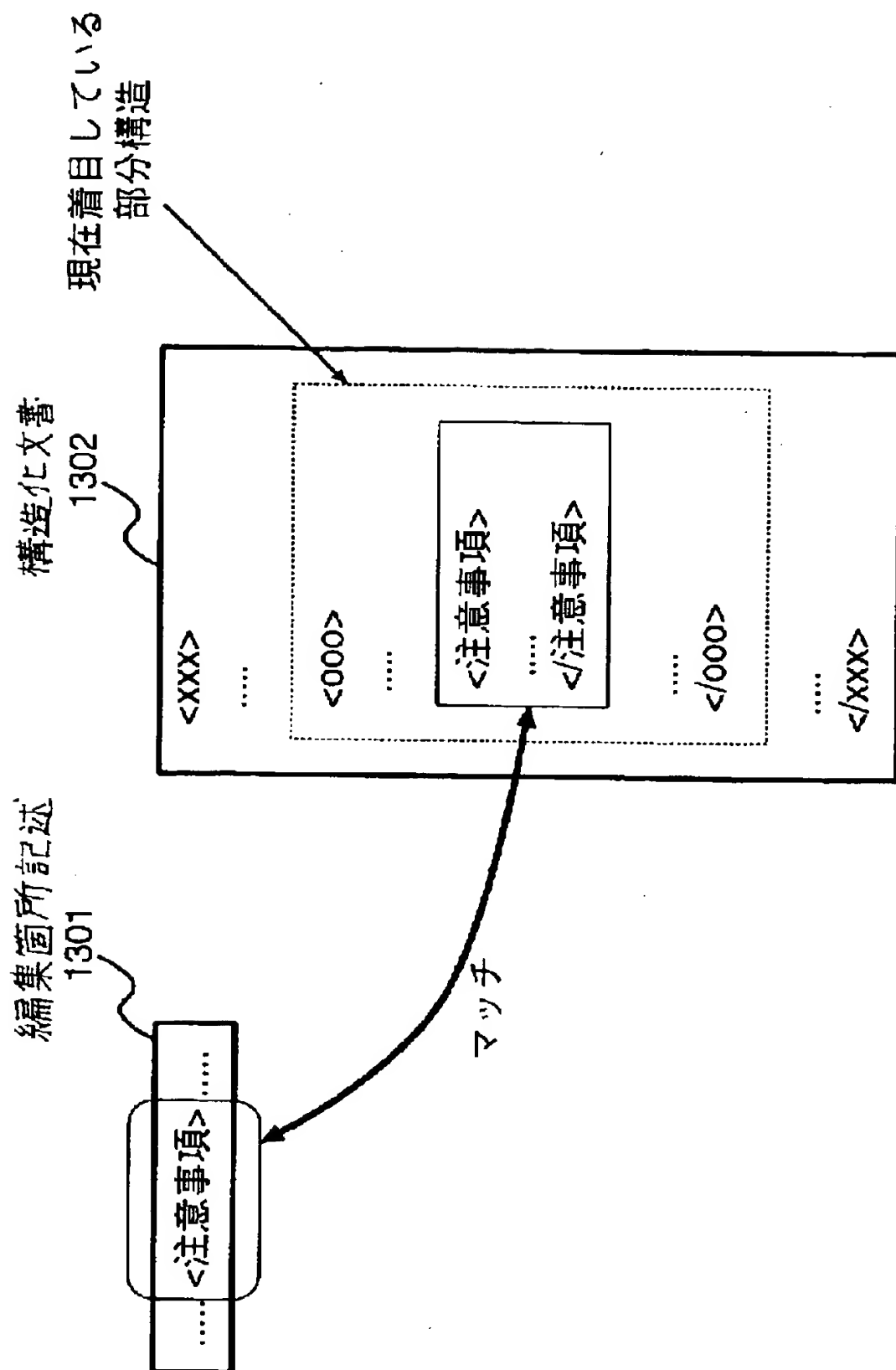
【図 1 1】

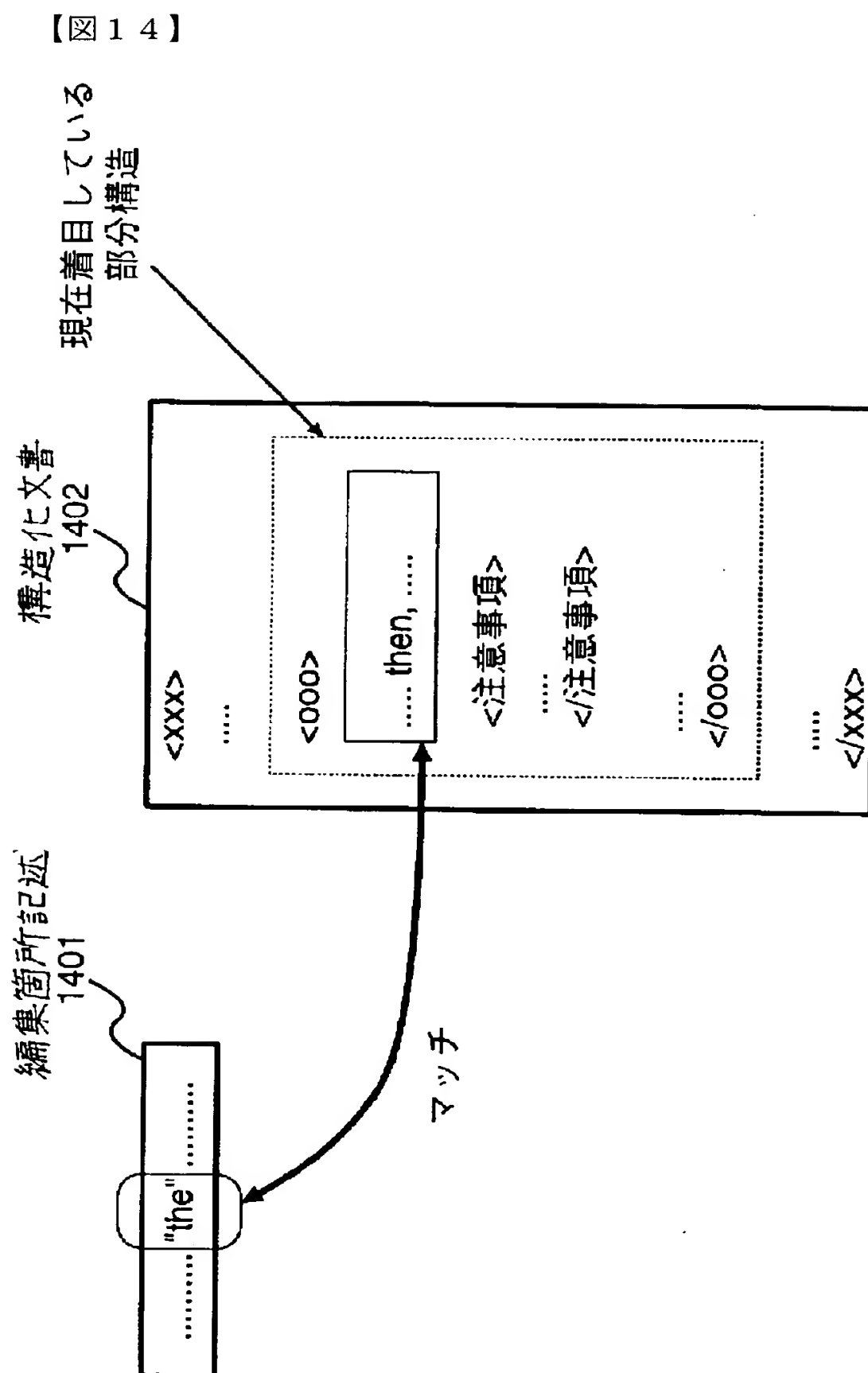


【図 1 2】

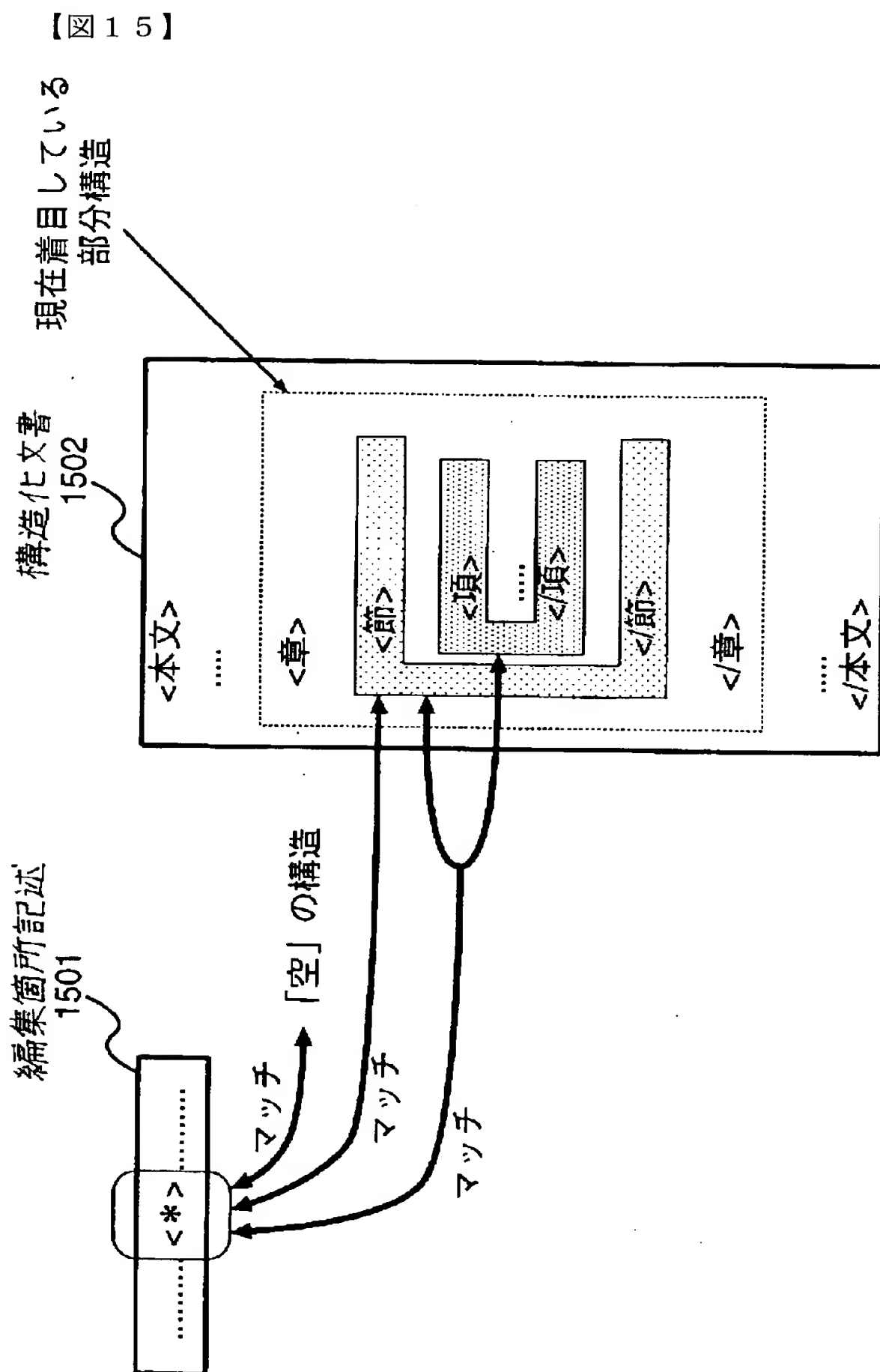


【図 13】

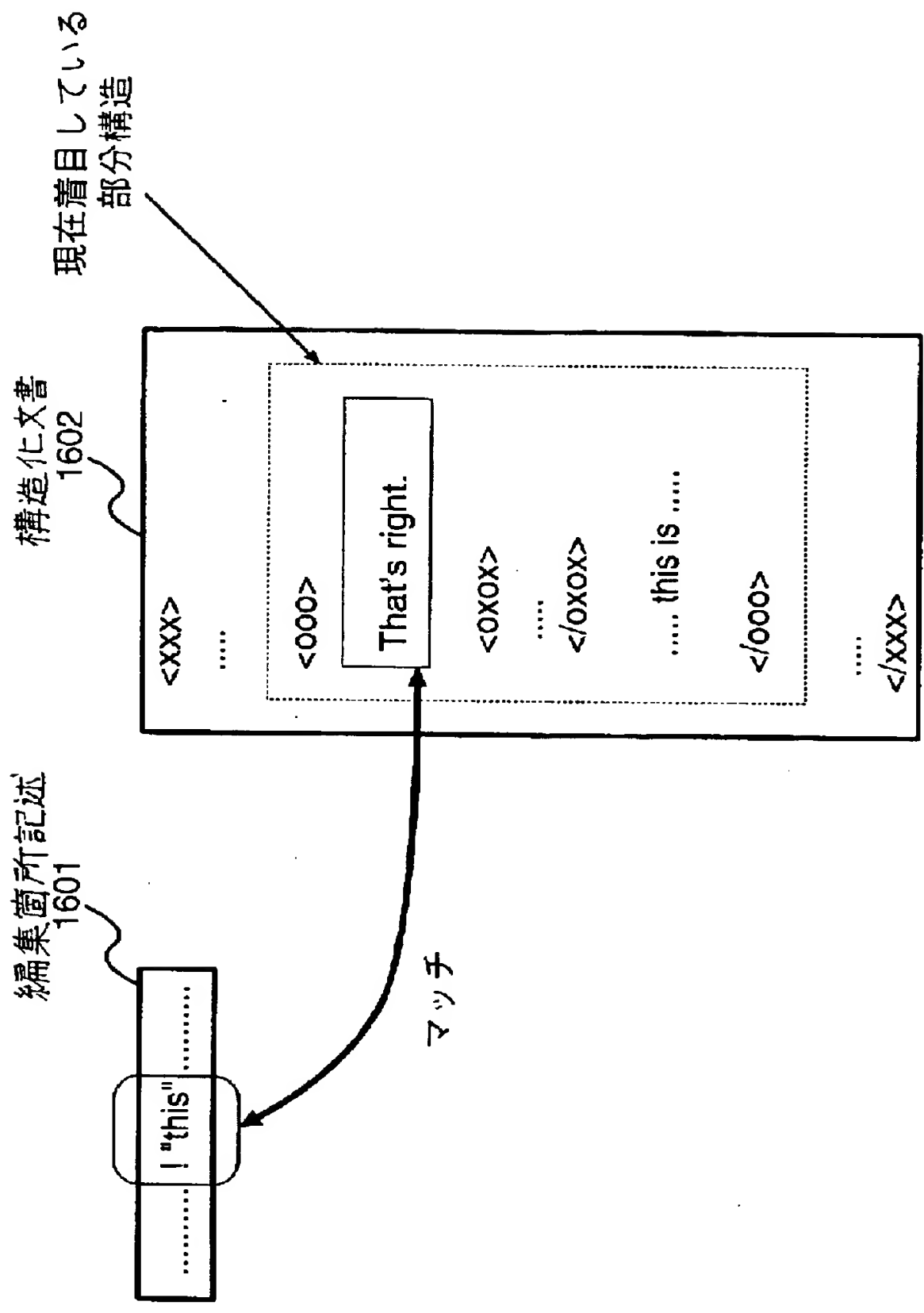




【図 1 4】

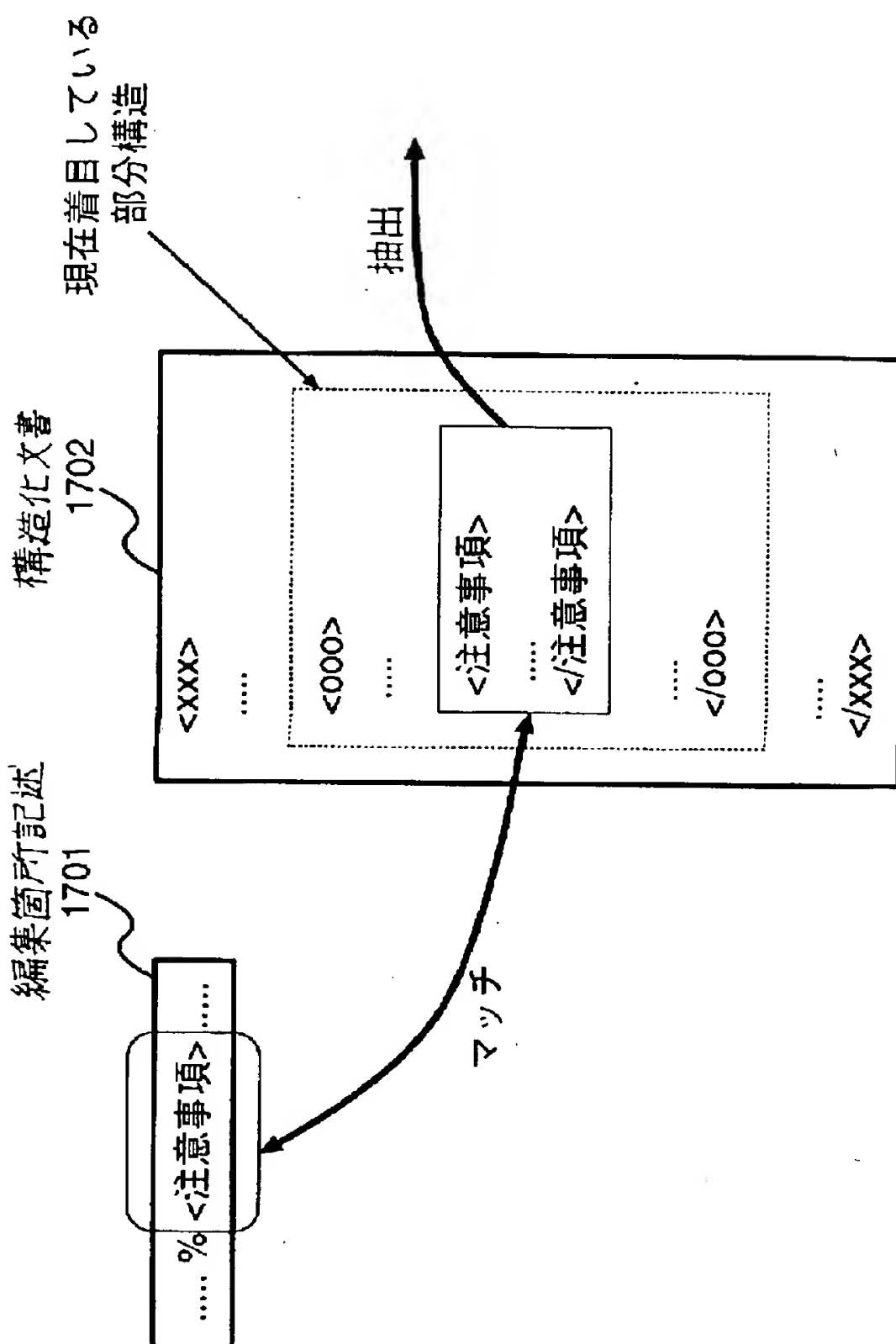


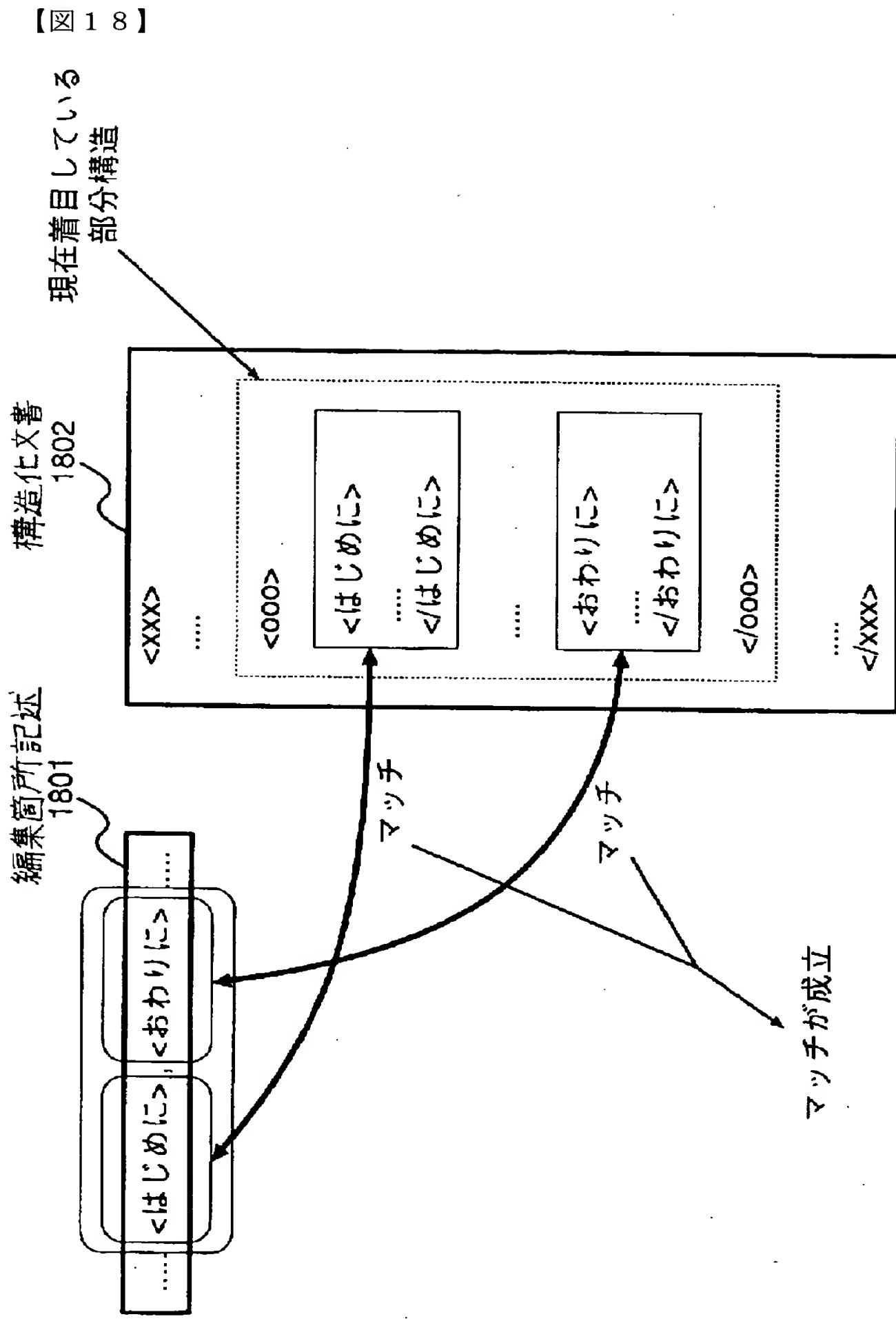
【図 15】



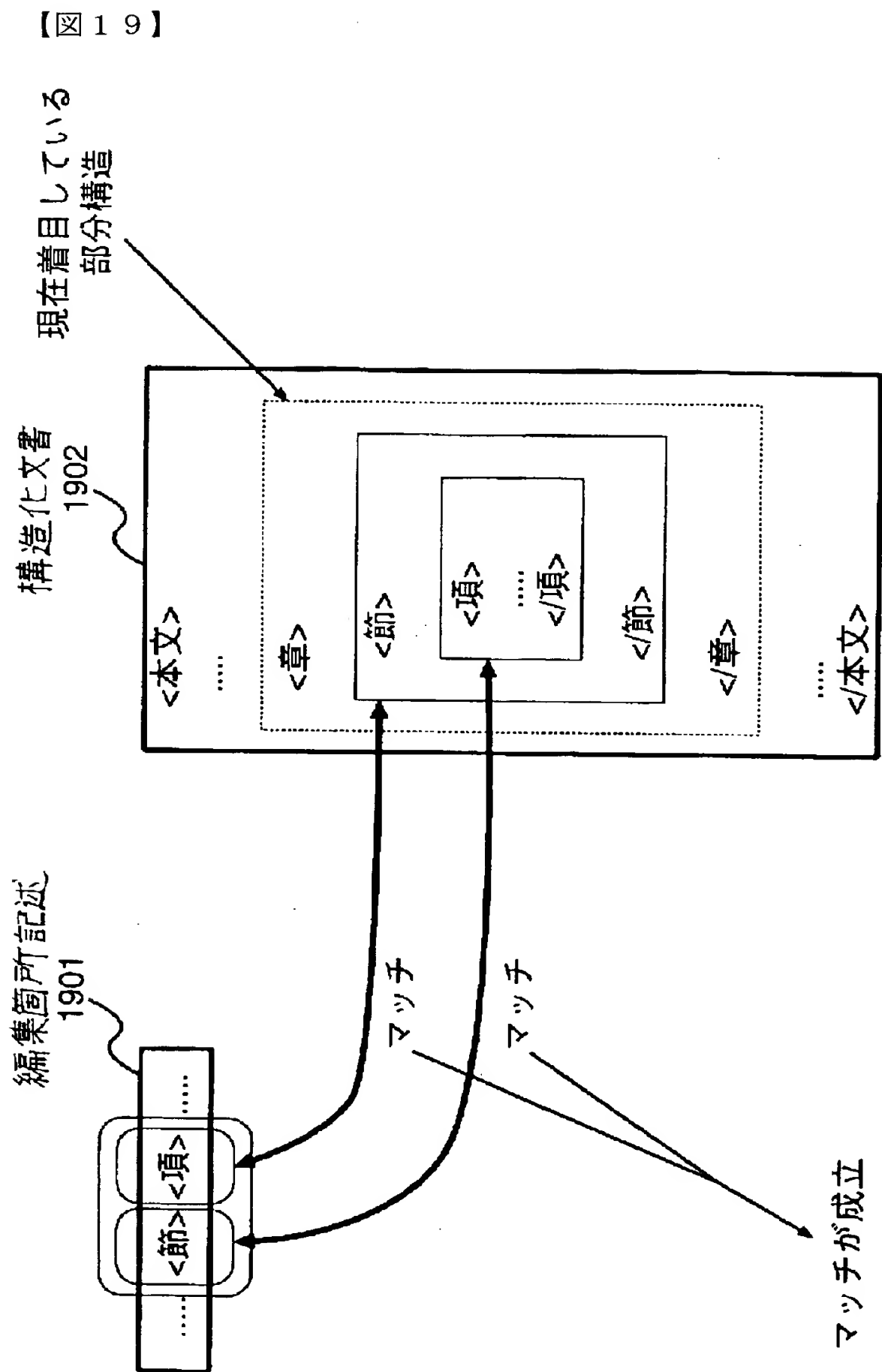
【図 16】

【図 17】

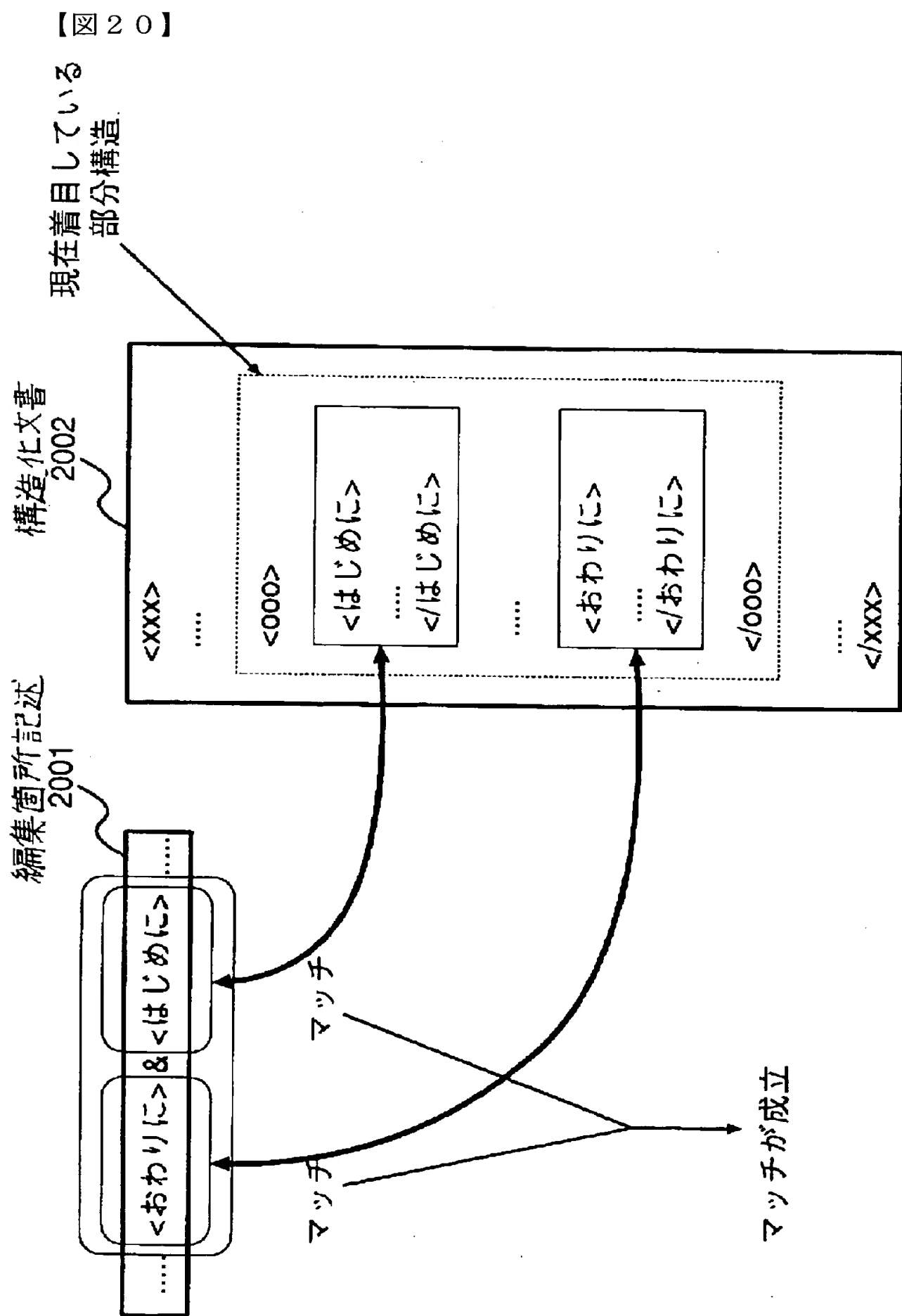


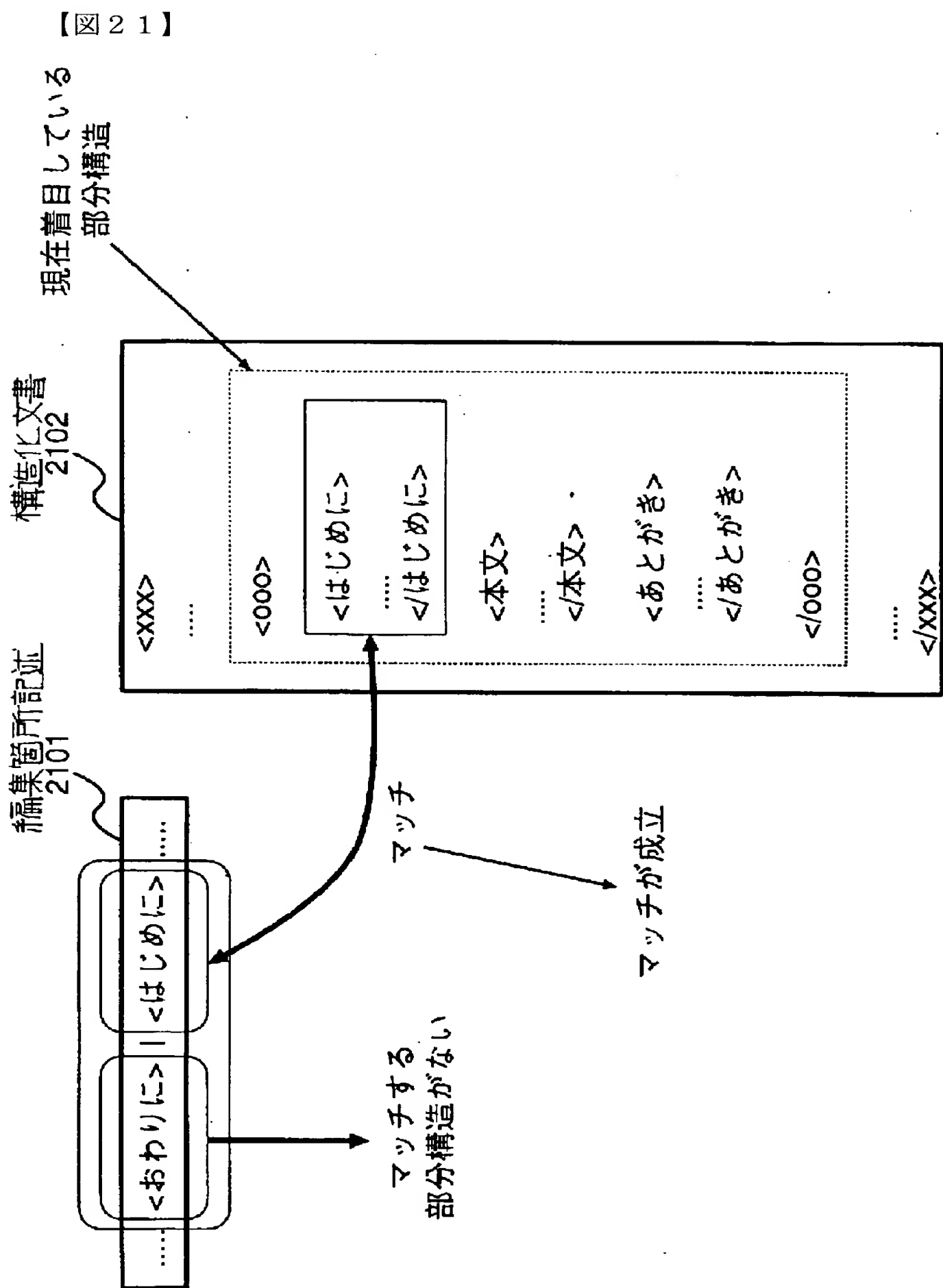


【図 18】



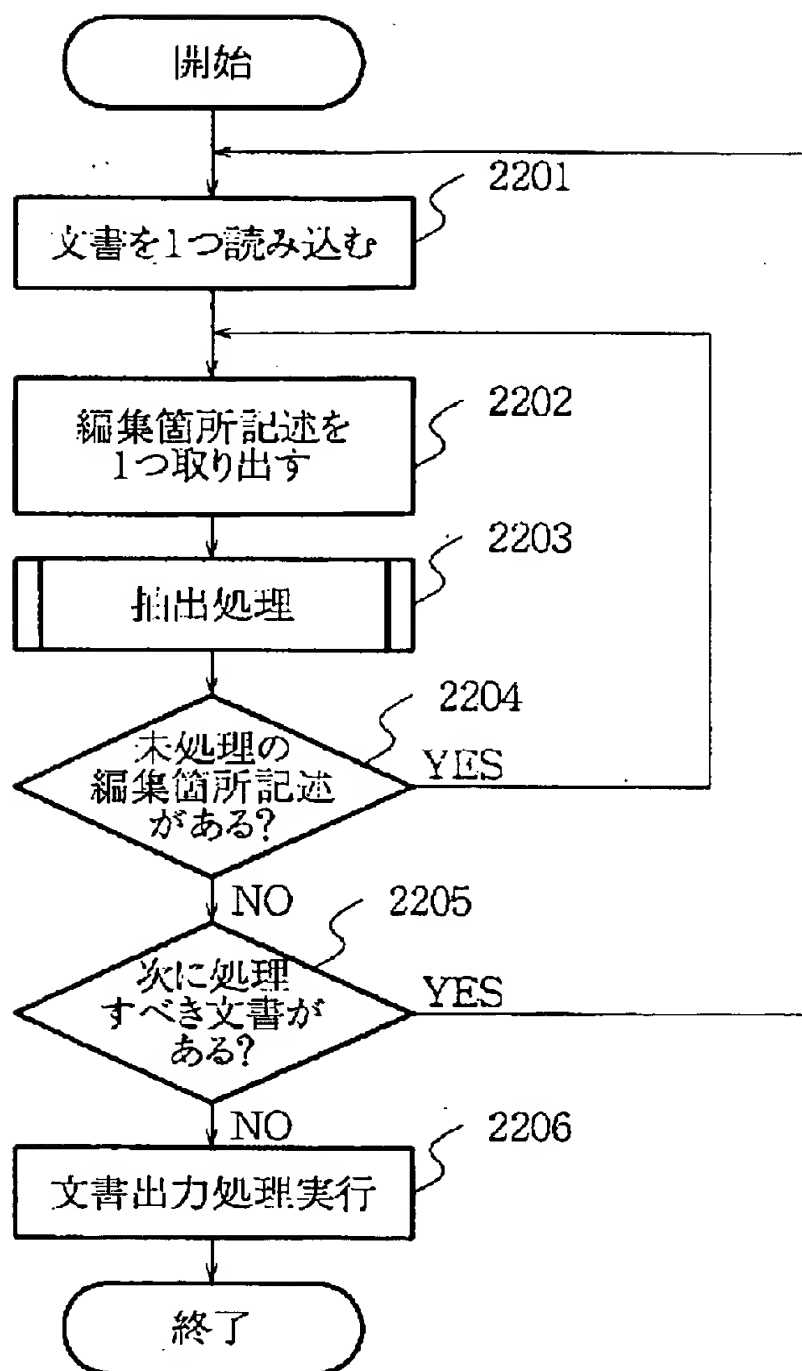
【図 19】



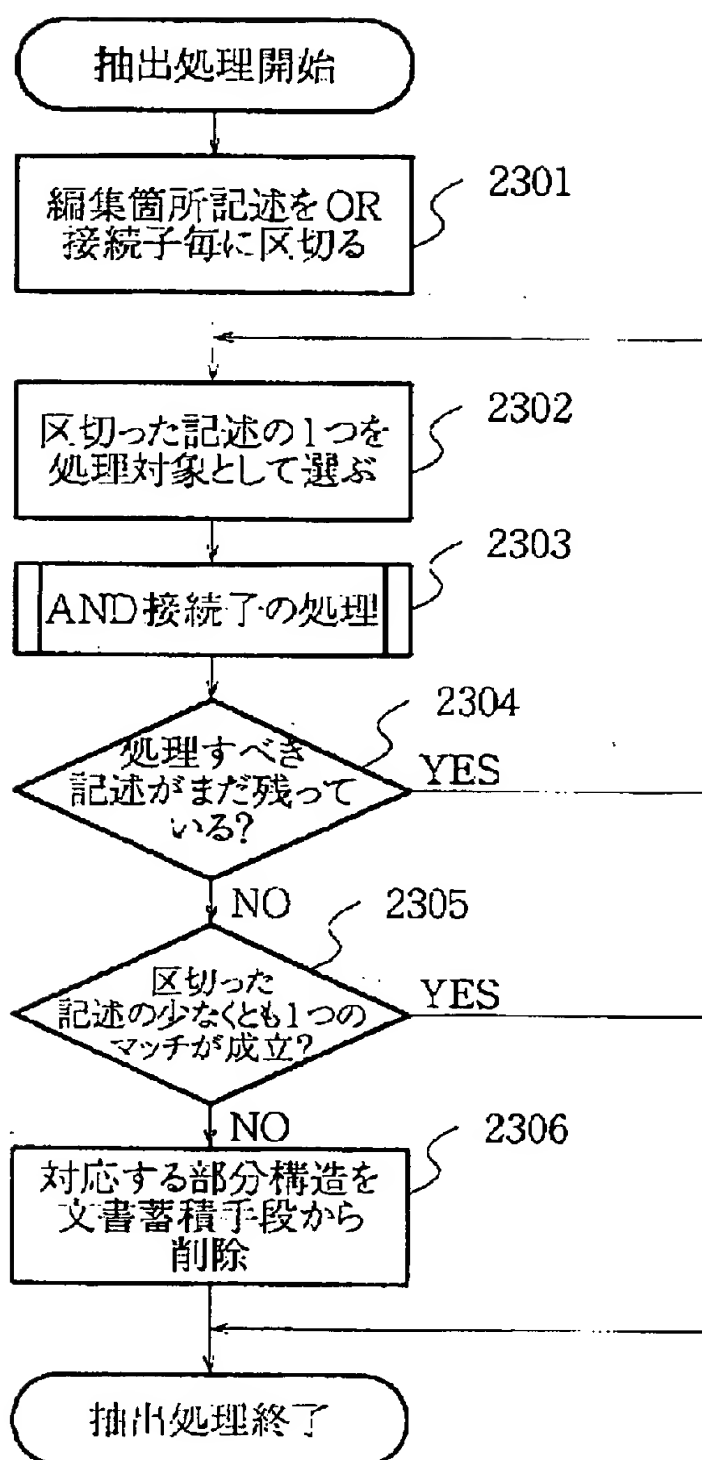


【図 2 1】

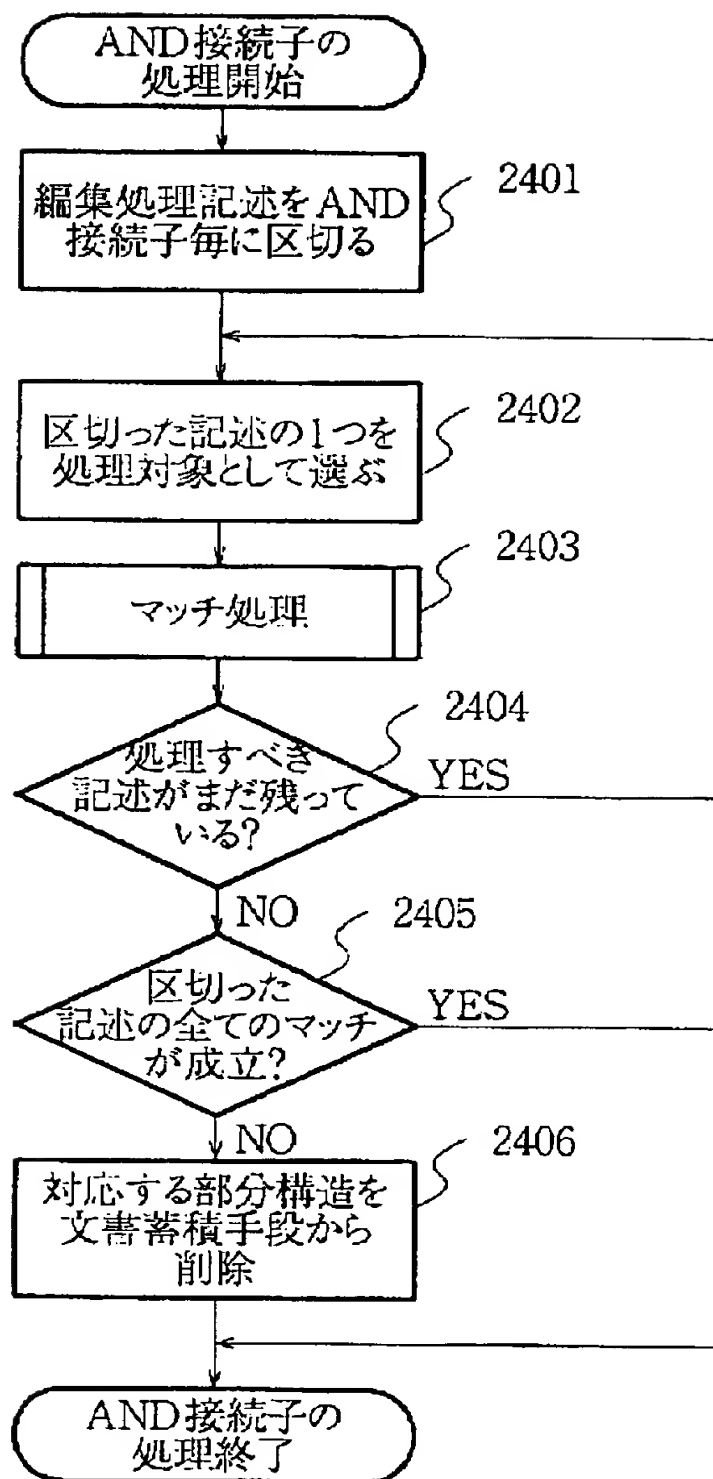
【図 2 2】



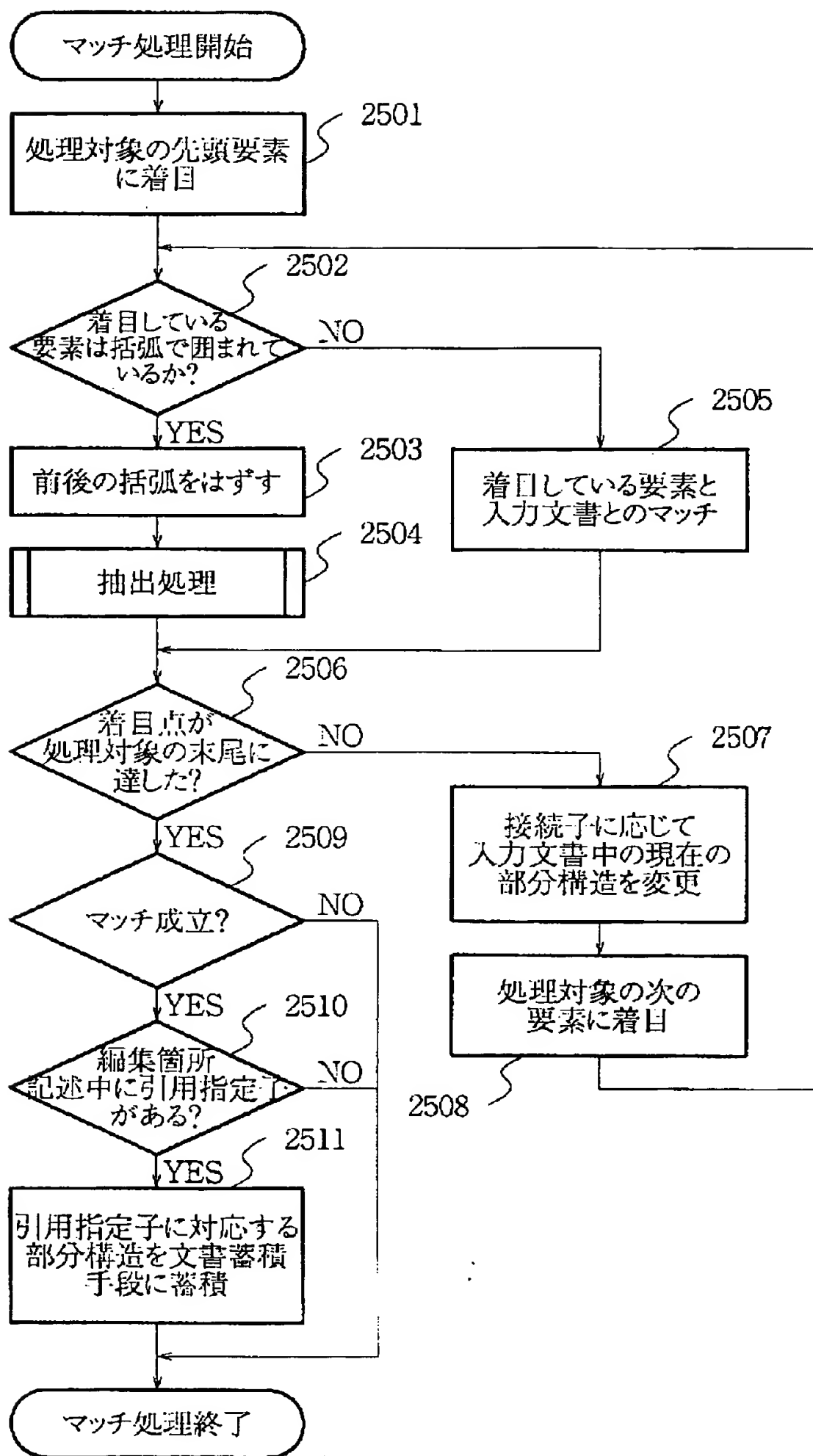
【図 2 3】

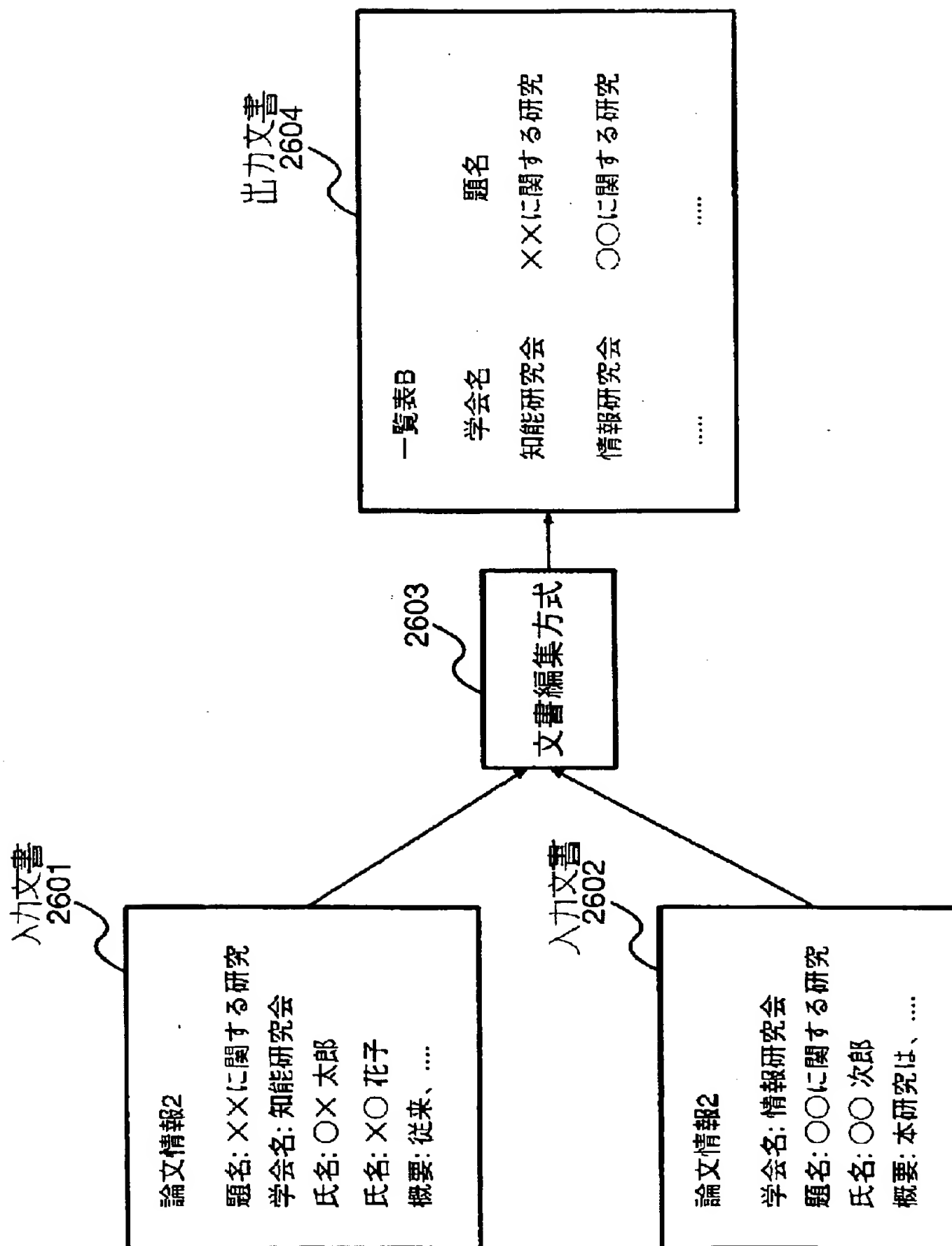


【図 2 4】

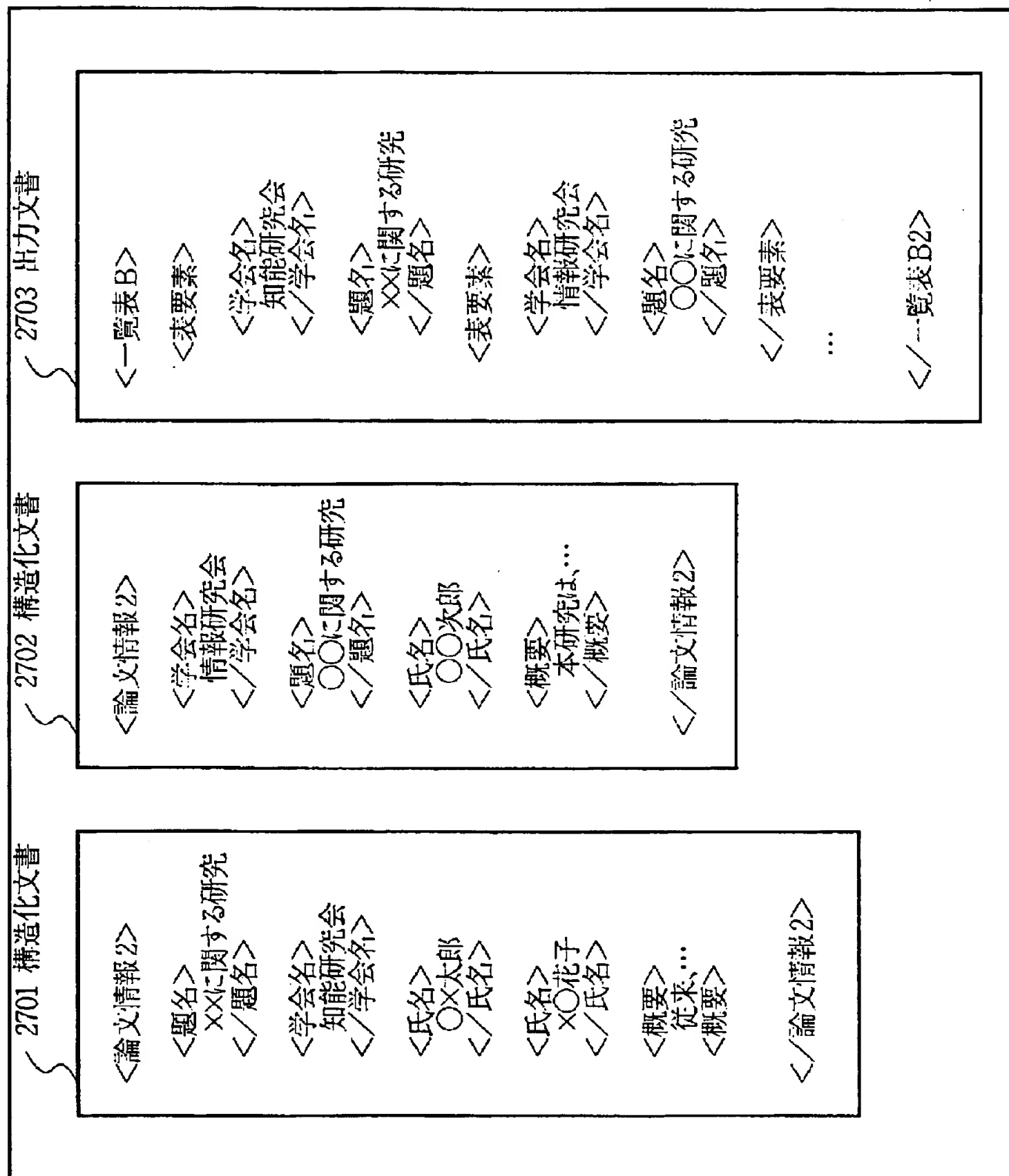


【図 2 5】





【図 2 6】



【図 2 8】

構造化文書  
2801

```
<論文>

<第一章>
  図1において、....

  <図>
    図1
  </図>

</第一章>

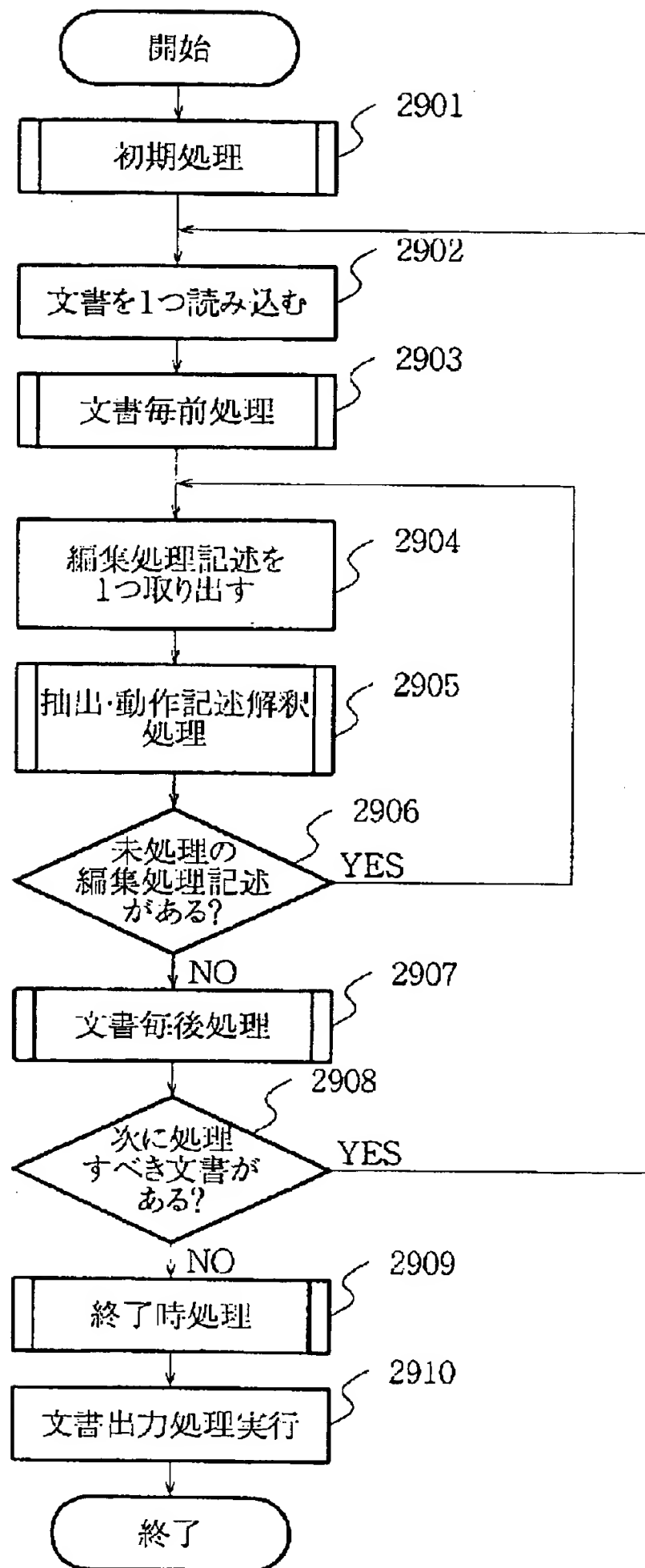
<第二章>
  図2において、....

  <図>
    図2
  </図>

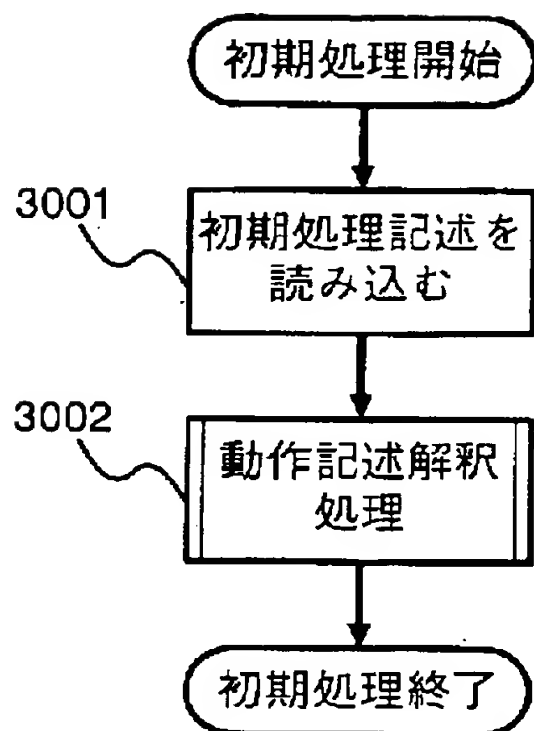
</第二章>

</論文>
```

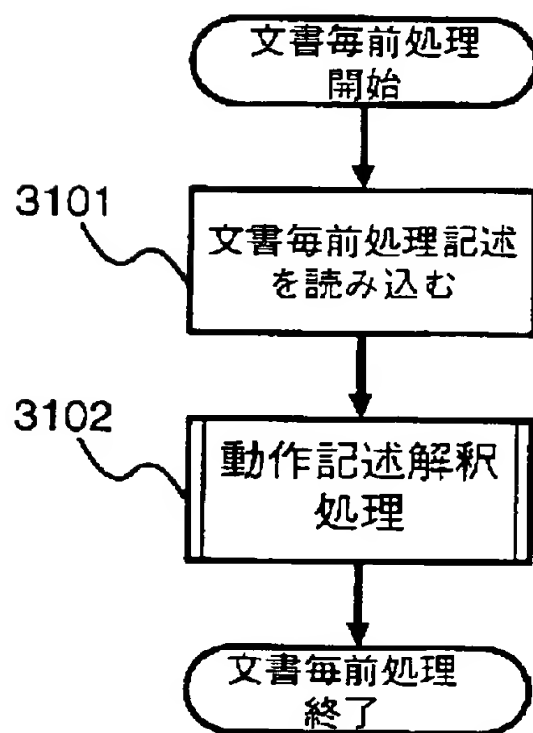
【図 2 9】



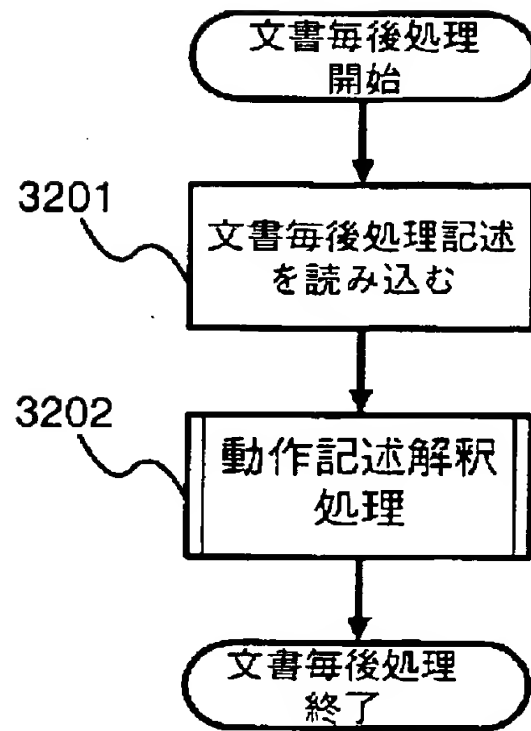
【図 3 0】



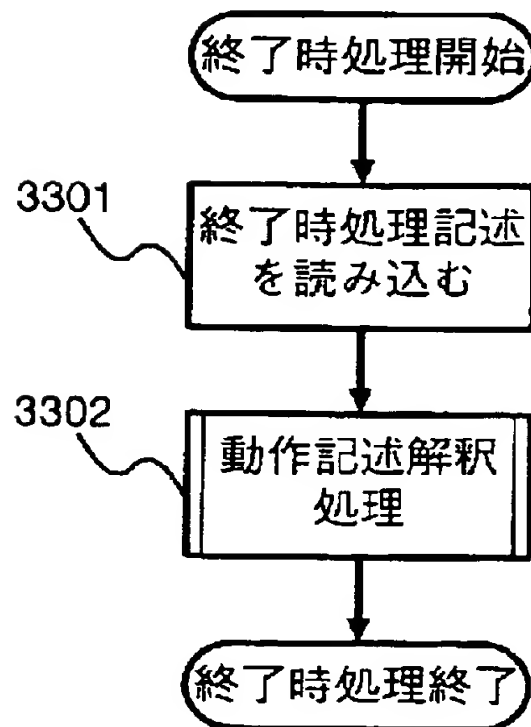
【図 3 1】



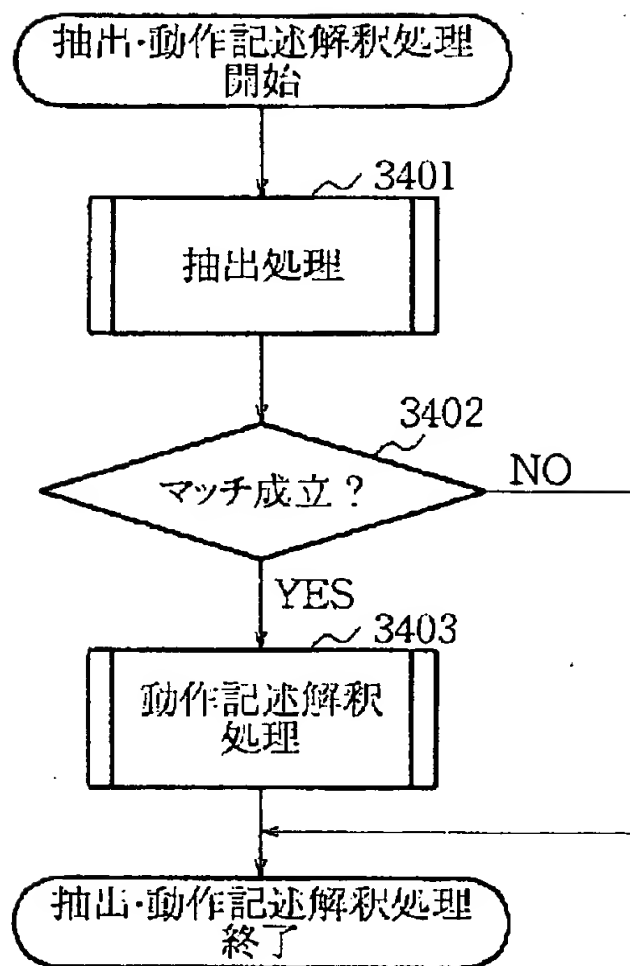
【図 3 2】



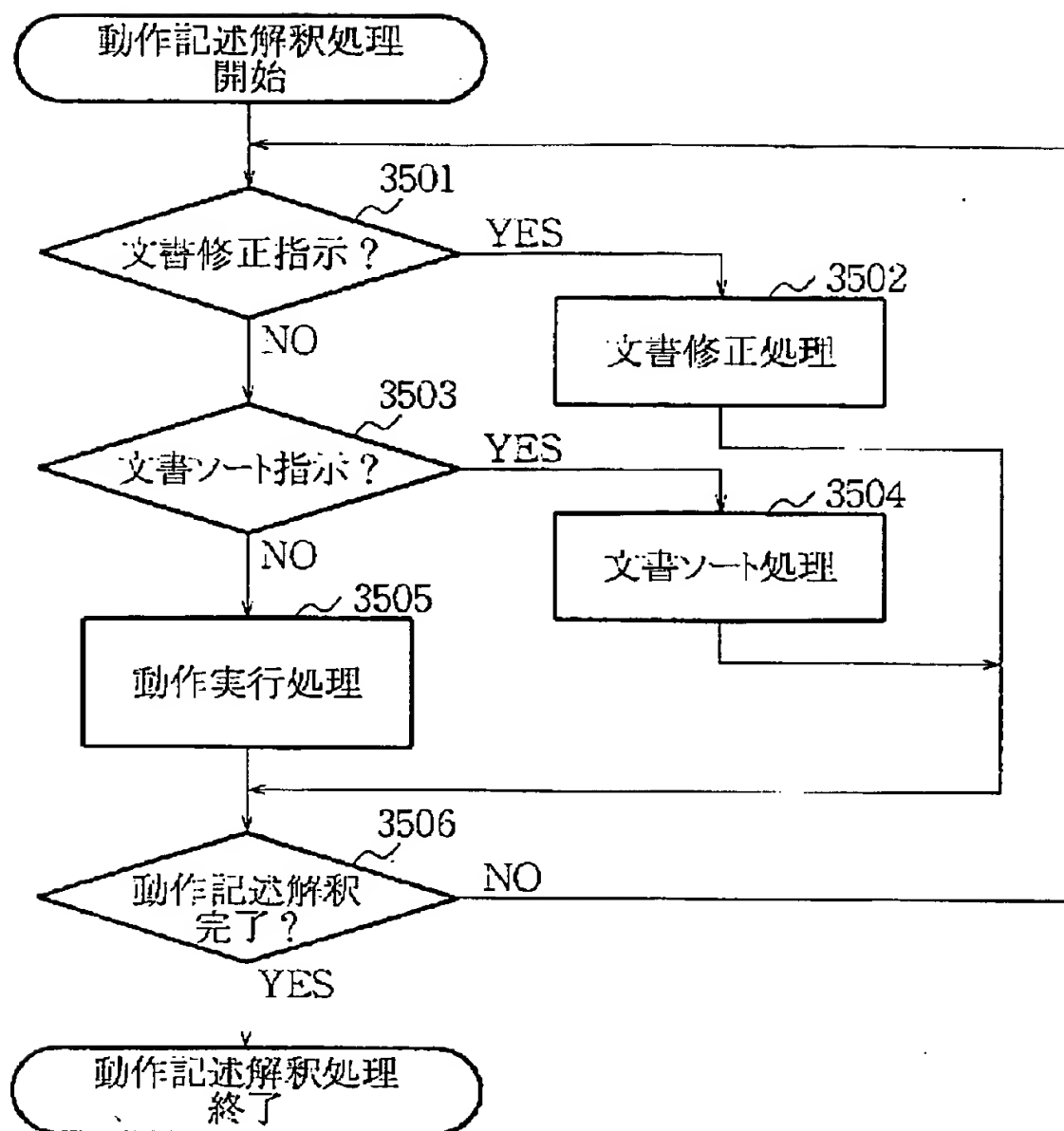
【図 3 3】



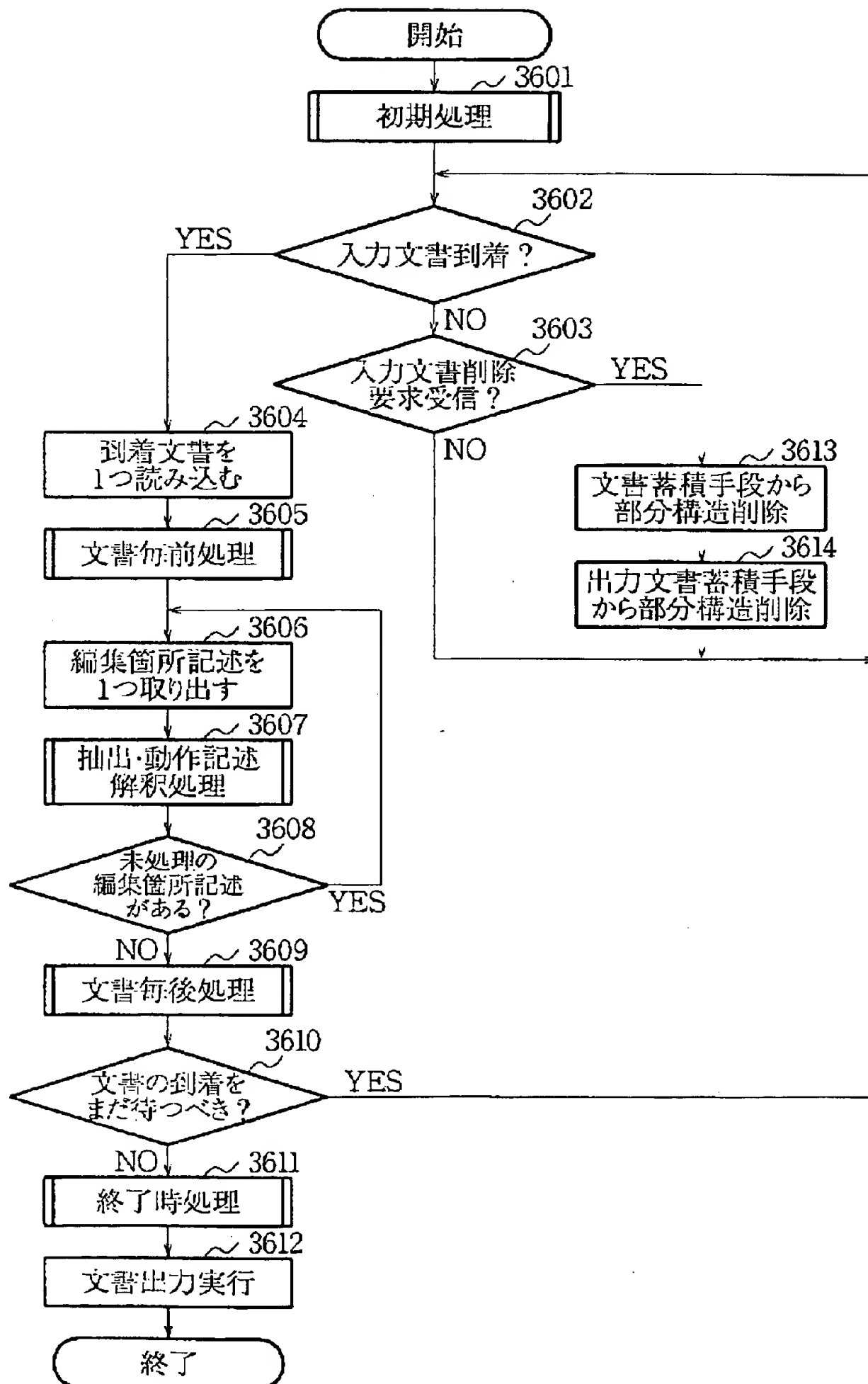
【図 3 4】

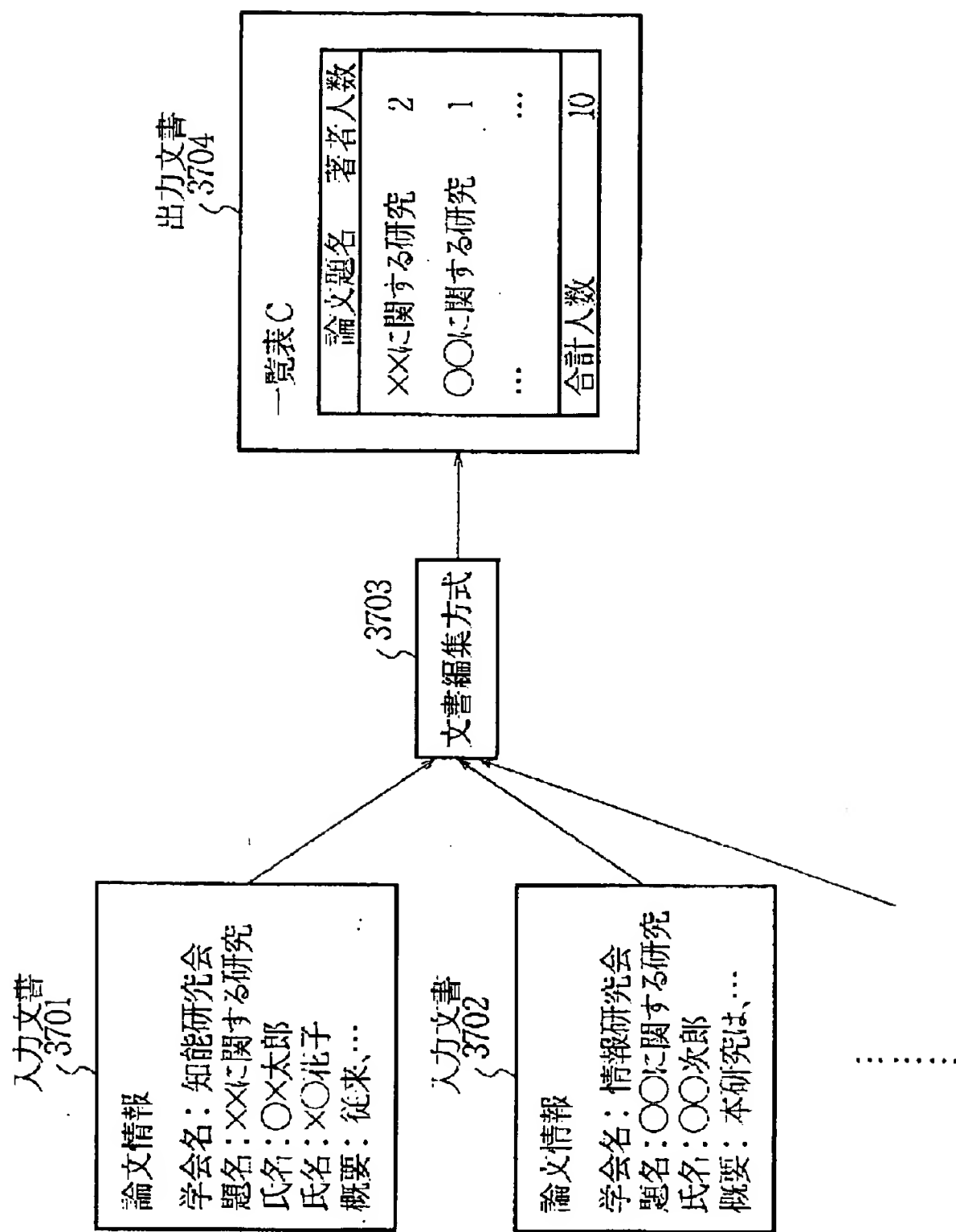


【図 3 5】



【図 3 6】





【図 3 7】

【図 3 8】

<一覧表C>	
<表要素>	
<論文題名>	XXに関する研究
</論文題名>	
<著者人数>	2
</著者人数>	
</表要素>	
<表要素>	
<論文題名>	〇〇に関する研究
</論文題名>	
<著者人数>	1
</著者人数>	
</表要素>	
.....	
<合計人数>	10
</合計人数>	
</一覧表C>	

【図 3 9】

構造化文書901が到着

(a)

一覧表C	
論文題名	著者人数
××に関する研究	2
合計人数	2

出力文書  
3901

構造化文書902が到着

(b)

一覧表C	
論文題名	著者人数
××に関する研究	2
〇〇に関する研究	1
合計人数	3

出力文書  
3902

構造化文書901の削除要求が到着

(c)

一覧表C	
論文題名	著者人数
〇〇に関する研究	1
合計人数	1

出力文書  
3903

【書類名】 要約書

【要約】

【課題】 複数の構造化文書から必要な部分構造を抽出してまとめ、新たな構造化文書を作成する場合などにおいて、文書の構造を単位として複数の抽出箇所を指定することのできるパターン記述に従って、抽出箇所間のつながりを維持しながら構造化文書から部分構造を抽出する。

【解決手段】 構造化文書を読み込む構造化文書入力手段101と、前記構造化文書に対する編集箇所記述が記述されている文書処理記述手段102と、前記構造化文書の部分構造を、前記編集箇所記述中の引用指定子毎に分け、かつ、前記構造化文書とのマッチが成立した順に蓄積する文書蓄積手段103と、前記編集箇所記述と構造化文書とのマッチ処理を行い、引用指定子で指定された部分構造を抽出して前記文書蓄積手段に蓄積する文書編集処理実行手段104と、前記文書蓄積手段に蓄積された文書をそのまま、あるいは加工して出力する文書出力手段105とを有する。

【選択図】 図 1